Advanced Simulation of Uncertain Quantities in Civil Engineering with Applications to Regional Hazards

Advanced Simulation of Uncertain Quantities in Civil Engineering with Applications to Regional Hazards

by

Vasileios Christou

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Doctor of Philosophy

in

Structural Engineering

Lehigh University

Bethlehem, PA

August 2019

Approved and recommended for acceptance as a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date

Dr. Paolo Bocchini
Dissertation Advisor

Accepted Date

Committee Members:

Dr. Shamim Pakzad
Committee Chairperson

Dr. Dan. M. Frangopol
Committee Member

Dr. Wei-Min Huang
External Member

Dr. George Deodatis
External Member

Dedication

To my parents, Μαρία and Μιχάλης

Acknowledgments

This would not have been possible without the unconditional support of my mentor, Dr. Paolo Bocchini. Thank you for investing your time, energy, wisdom, love, and compassion into my development. Thank you for the many opportunities you have provided me to grow professionally and for the example you provide as a researcher, scholar, teacher and human.

Thank you to the members of my thesis committee, Dr. Shamim Pakzad, Dr. Dan M. Frangopol, Dr. Wei-Min Huang and Dr. George Deodatis for your continued support and for taking the time to review my work.

Thank you to Dr. Daniel Conus for your time, teaching and help during the research process.

The support from

- the United States Army Contracting Command Picatinny Arsenal (ARDEC) through the Pennsylvania Innovation and Advanced Technology Association (PIATA), under project "Integrated Framework for the Application of Probabilistic Optimization Technology (POTech) to Weapons with Emphasis on Modeling and Simulation";
- the National Science Foundation through grant CMS-1541177 "Probabilistic Resilience Assessment of Interdependent Systems (PRAISys)" www.praisys.org, part of the CRISP program and the CMMI Division;
- the Pennsylvania Department of Community & Economic Development through grant PIT-16-12;

- the Department of Civil and Environmental Engineering and the P.C. Rossin College of Engineering and Applied Science, through Teaching Assistantship and other fellowships

is gratefully acknowledged. The opinions and conclusions presented in this thesis are those of the author and do not necessarily reflect the views of the sponsoring organizations.

Table of Contents

List of Tables

List of Figures

Abstract

During a natural extreme event, such as an earthquake or a hurricane, the amount of socioeconomic losses due to inefficient disaster response or the losses due to long-term reduction of functionality of infrastructure systems is comparable, if not higher, to the immediate losses due to the extreme event itself. Therefore, the scientific community has recognized the need to be able to accurately predict the performance of the lifelines and infrastructure when a extreme event occurs and also being able to build structures and infrastructures which are able to successfully withstand extreme events, effectively satisfy the needs of the post-event emergency response, and restore its functionality as soon as possible.

One of the first challenges in performing such tasks is to accurately assess the load of an extreme event to the geographic region the infrastructure or lifeline belongs in. Hazard maps and probability of exceedance curves are very popular tools used initially for the probabilistic seismic hazard analysis and expanded later to other hazards such as hurricanes. These tools provide the probability of exceeding any given value of an Intensity Measure (IM) of choice (e.g., 1-minute sustained wind speed) at any location. These tools are an integral part of the performance-based design approach and are essential for the probabilistic analysis of individual structures. However, these tools are not appropriate for the analysis of distributed infrastructure systems because they do not account for the correlation information among the values of the IM at different locations. Engineers have recognized that the various network components cannot be studied independently because the performance of the entire network depends on the combination of the conditions of all members. Thus, considering joint probabilities of having certain values of the IM at the

locations of interest if required.

The most widely accepted approach by the scientific community to address these issues is through simulation-based techniques. Based on this idea, a set of representative extreme scenarios is selected and then the impact on the network components and, in turn, the performance of the network itself is predicted. The main issue in this approach is the computational cost, which constraints the number of extreme event scenarios to be as small as possible, while the set still captures the probabilistic characteristics of the intensity of the investigated natural extreme event over a region.

A new framework is presented for the selection of an optimal set of stochastic intensity measure maps representing the regional hazard over a geographic area. This set of IM maps can subsequently be used for the analysis of spatially distributed infrastructure systems. The proposed methodology results in a versatile multihazard tool that accounts for the spatial correlation through the optimal sampling of IM maps. Its key characteristic is that it embraces the nature of the regional IM maps as two-dimensional random fields. The representation of the regional hazard is supported by proofs of optimality, ensuring mean-square convergence of the ensemble of representative IM maps to the complete portfolio of possible hazard events, which is a particularly important property for risk analysis. A detailed comparison of the proposed technique with other popular methodologies in the same filed is presented.

Before applying the proposed technique or any other hazard representation technique, it is necessary to accurately study and characterize the regional hazard in a probabilistic way. Two types of natural phenomena were considered in the conducted research for the regional hazard analysis: the earthquake and the hurricane hazard. In the case of earthquakes, the seismic characterization of the Charleston South Carolina region was studied and a seismic modeling procedure was developed which includes spatial and temporal in-

formation and descriptions of fault geometry and style as well as other parameters. With the complete probabilistic description of the regional seismic hazard, the ground-motion prediction equations (GMPE) are implemented. The GMPE account for the in between-earthquake and within-earthquake variability, and result the ground shaking acceleration over the region. In the case of hurricanes, a more holistic approach was adopted by stochastically modeling the hurricane's track. Historical hurricane events, originated either in the Atlantic basin, Caribbean Sea or the Gulf of Mexico, have shown to significantly affect geographic regions located in the South and Eastern U.S. Therefore, a simulation framework is developed for the prediction of hurricane wind intensity and direction over any geographic region in the Southern and East U.S. The proposed framework generates synthetic hurricane directional wind speeds, and does so by utilizing historical data, simulating the hurricane's track and intensity, simulating key characteristic parameters such as the central pressure and the radius to maximum wind among others both for offshore and overland locations of the track, performing a wind field analysis, calculating the 10-meter wind intensity by utilizing oceanic- and land-based boundary layer models and finally simulating offshore and overland wind directions.

Chapter 1

Introduction

1.1 Motivation

Severe natural disasters occurred during the past few years have caused the scientists to focus their attention on the effects of the natural disaster on the community level. For instance, the 2011 Tohoku earthquake and tsunami in Japan on March 11, 2011 (Takewaki, 2011) caused approximately $319 billion in direct losses and more than $619 billion in indirect losses (Tesfamariam and Goda, 2013). The 1987 Ecuador earthquake had up to 7 times greater indirect losses than direct losses due to business interruption and lost revenue (Daniell et al., 2012). The 1994 Northridge earthquake caused over $1.5 billion in business interruption losses because of damage to the transportation network (Chang, 2003). The 2009 L'Aquila earthquake in Italy showed comprehensive emergency response in the short term but poor performance on the long-term reconstruction phase (Cimellaro et al., 2010). The high indirect losses and the socio-economic disruptions after the occurrence of an extreme event show that the built environment needs to be studied in its entirety in order to further mitigate natural disaster effects and to more accurately forecast the effects of extreme natural events.

Much attention has been devoted to studying and developing resistant structural components and systems in order to reduce the direct losses after a natural disaster (e.g. earthquake). For this reason, probabilistic hazard analysis methods to address the seismic hazard were developed for site-specific locations (Cornell, 1968). This later became known

as probabilistic seismic hazard analysis or PSHA (McGuire, 2004). PSHA focuses on the evaluation of the likelihood of strong ground motions and the main outcomes are the seismic hazard curve (i.e., graphical plot depicting the annual frequency of exceedance vs ground motion severity) and the uniform hazard spectrum (i.e., expected ground motion levels for a given annual frequency of exceedence vs structural period) (McGuire, 2008). These essential tools were the basis of modern seismic design provisions in building codes, and they formed the ground to develop the seismic hazard maps.

The use of the seismic hazard maps and the seismic hazard curves is an integral part of the probabilistic seismic risk analysis (PSRA). Furthermore, hazard maps and hazard curves are an essential part of the performance-based earthquake engineering framework or PBEE (Cornell et al., 2002; McGuire, 2004; Ruiz-Garc´ı a and Miranda, 2007), the use of which is essential for the probabilistic analysis of individual structures. The Pacific Earthquake Engineering Research (PEER) PBEE framework is summarized in Figure 1.1 and can be described in terms of four main analysis steps. These separate —but related— phases comprise the *hazard analysis*, which results in the seismic characterization of a geographic location; the *structural analysis*, which provides the response of a structure as a function of deformations or forces; the *damage analysis*, which yields a description of the damage of the structural and non-structural elements; and, lastly, the *loss analysis*, which translates the damage into losses based on certain measures of interest. Mathematically, the outcome of each step can be characterized by generalized variables: *Intensity Measure (IM), Engineering Demand Parameter (EDP), Damage Measure (DM) and Design Variable (DV)*, respectively. The goal of the PEER framework is to enhance the accuracy of the results used in decision making by considering the uncertainty in a rigorous probabilistic manner.

Each step builds upon the previous step, and all four steps depend on initial location information and the facility's structural characteristics. In seismic engineering, the variable

of the *hazard analysis* phase (i.e., *IMs*) denotes the ground motion intensity (e.g. spectral acceleration, peak ground acceleration) at a certain location and is based on some characteristic period of the structure, accounting for the uncertainty of earthquake magnitude-recurrence rates, epicenter location etc. Mathematically, this can be expressed with the term $g(IM|D)$. In the *structural analysis* phase, the engineer estimates the structural response of the facility considering measures in terms of the EDP, conditional on the ground motion characteristics and geographic location (i.e., $p[EDP|IM,D]$). Structural responses include, but are not limited to, the roof displacement, the inter-storey drift or the internal member forces. This step accounts for the uncertainty in material properties, damping, force-deformation characteristics of the model, etc. The *Damage analysis* then uses values of the EDP variable as an input parameter and yields fragility functions which computes the probability that the structure exceeds some undesirable limit state conditional on the facility response (i.e., $p[DM|EDP,D]$). This step considers the uncertainty related to the structural members capacity and the load history, among others. The last stage in the PEER process is the *Loss analysis*, which estimates the asset performance parameterized by a DV, such as monetary value, down-time, etc. ($p[DV|DM,D]$). Keeping each step in mind, the methodology can be expressed in terms of a triple integral utilizing the total probability theorem

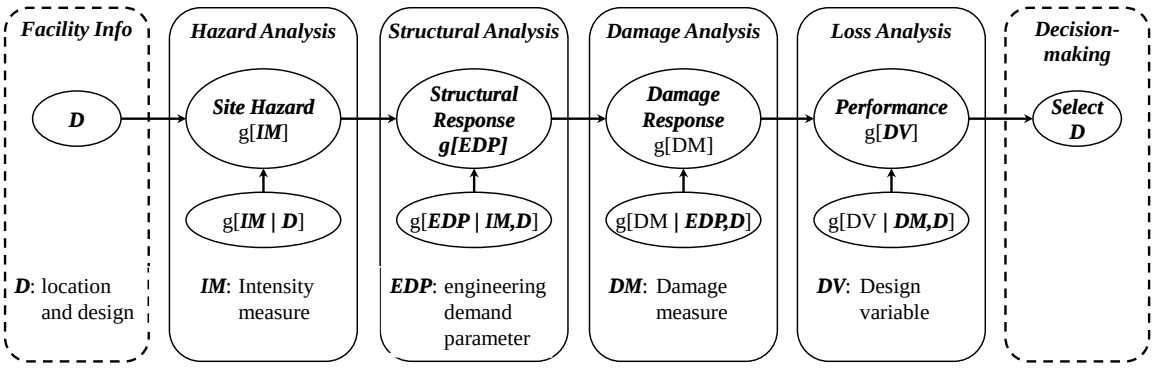


Figure 1.1: PEER performance-based loss analysis framework (Porter, 2003).

(Porter, 2003):

$$g[DV|D] = \int\int\int p[DV|DM,D]\, p[DM|EDP,D]\, p[EDP|IM,D] \\ g[IM|D]\, dIM\, dEDP\, dDM \tag{1.1}$$

Equation (1.1) when applied to a single-site, can be solved either numerically or using simulation-based methods. The former approach is adopted when all the functions describing the uncertainties in each stage of the analysis and the inter-relationships of subsequent phases are defined in closed-form expressions. The latter, more general, approach was first applied to PSHA (Ebel and Kafka, 1999; Musson, 1999, 2000), and then to the rest of the PBEE framework. The available tools resulting from the PSHA analysis phase and the PBEE framework, however, are not in general, easily applicable to engineering problems taking into account distributed infrastructure systems, lifelines and building stocks. There are four main reasons which make them inappropriate, as described below.

The first reason is that the tools yielded by the PSHA stage (e.g., hazard maps) do not provide information on the correlation of the measures used to quantify the intensity of the hazard (IM) between spatially distributed assets (see review papers by Sokolov and Wenzel, 2011b,a). Instead, when a spatially distributed system is considered, it is necessary to know the probability of having *simultaneously* certain values of the IM at all locations of interest (Bocchini and Frangopol, 2011a, 2012; Decò et al., 2013; Saydam et al., 2013). Another reason that makes the resultant tools impractical is that for various sites, the scalar intensity measure in Equation (1.1) should be replaced with a vector of IMs, which would add another layer of complexity to the integral. Furthermore, the variability on each step conditional to the previous steps forces each step to be repeated many times in a simulation pattern, which results in high computational demands that are infeasible given the commonly available computational resources. Lastly, the connection between the damage response and the performance of a spatially distributed system is usually not available in

closed form solution (i.e., $p[DV|DM,D]$). For example, the optimal bridge restoration sequence of a transportation network for a given damage state of network bridges after the occurrence of an extreme event is not available in closed-form. Problems of this type are usually solved utilizing numerical optimization techniques (Karamlou and Bocchini, 2014).

The research presented in this thesis addresses the challenges involved in a regional probabilistic hazard analysis and developed a comprehensive methodology for the accurate probabilistic quantification of the regional hazard which can effectively be used by the subsequent steps of a regional loss estimation analysis or a network risk assessment. A new, versatile, and effective multi-hazard simulation-based approach is the main goal of the conducted research applied to natural hazards such as earthquakes and hurricanes.

The proposed approach consistently addresses the investigated hazards as random fields. The goal of this methodology is to optimally sample a limited number of IM maps while properly considering the spatial correlation and the uncertainties involved.

1.2 Research overview

Simulation-based regional probabilistic hazard analysis requires the development of simulators which can generate synthetic hazard scenarios that accurately capture the extreme event's probabilistic characteristics. These simulators should account for the extreme event's recurrence rates and its intensity variability and uncertainty over space, among other features. In addition, the simulators should yield the extreme event's resulting loads on the built environment, which can be univariate or multivariate. For example, the peak ground acceleration in the case of an earthquake scenario can be interpreted as a univariate load, or the maximum wind intensity and directionality in the case of a hurricane scenario can be interpreted as a two variate load.

In regional hazard analysis —which is the focus of this research, as discussed in Section 1.1— the information on the spatial distribution of the load over a geographic region is essential. The resulting loads over a geographic region are characterized by uncertainty and correlation, which makes them multidimensional and non-Gaussian random fields. Such simulators, then, can be very complex–both scientifically and computationally.

In order to acquire the skills for the simulator's development, the author was exposed to various research tasks and projects that built a strong foundation for this simulator's development. Each research project and task contributed to the development of the multi-hazard simulator by allowing the researcher to study and develop new techniques to simulate random variables and functions that had distinct probabilistic properties. A thorough chronological overview of the research tasks follows, with an accompanying pictorial representation (Figure 1.2). The topics presented in the blue signs in Figure 1.2 are briefly described in this chapter; however, they are not presented in details in subsequent chapters. On the other hand, the tasks presented within the green signs are described in this chapter and subsequently discussed in details in the remaining chapters of this thesis.

One of the most lengthy tasks (4-years duration) is related with the research project entitled "**Integrated framework for the application of probabilistic optimization technology (POTech) to weapons with emphasis on modeling and simulation**." The objective of this collaborative task among the author and Professors Frangopol, Bocchini, and Sabatino was to probabilistically assess the life-cycle of weapon systems for the U.S. Department of Defense - Research, Development, and Engineering Command - Armament Research, Development and Engineering Center (REDCOM-ARDEC). Throughout the project's course, the researcher performed the following tasks:

- Investigated failure modes of a weapon system
- Investigated relationship between events, malfunctions, and physical components

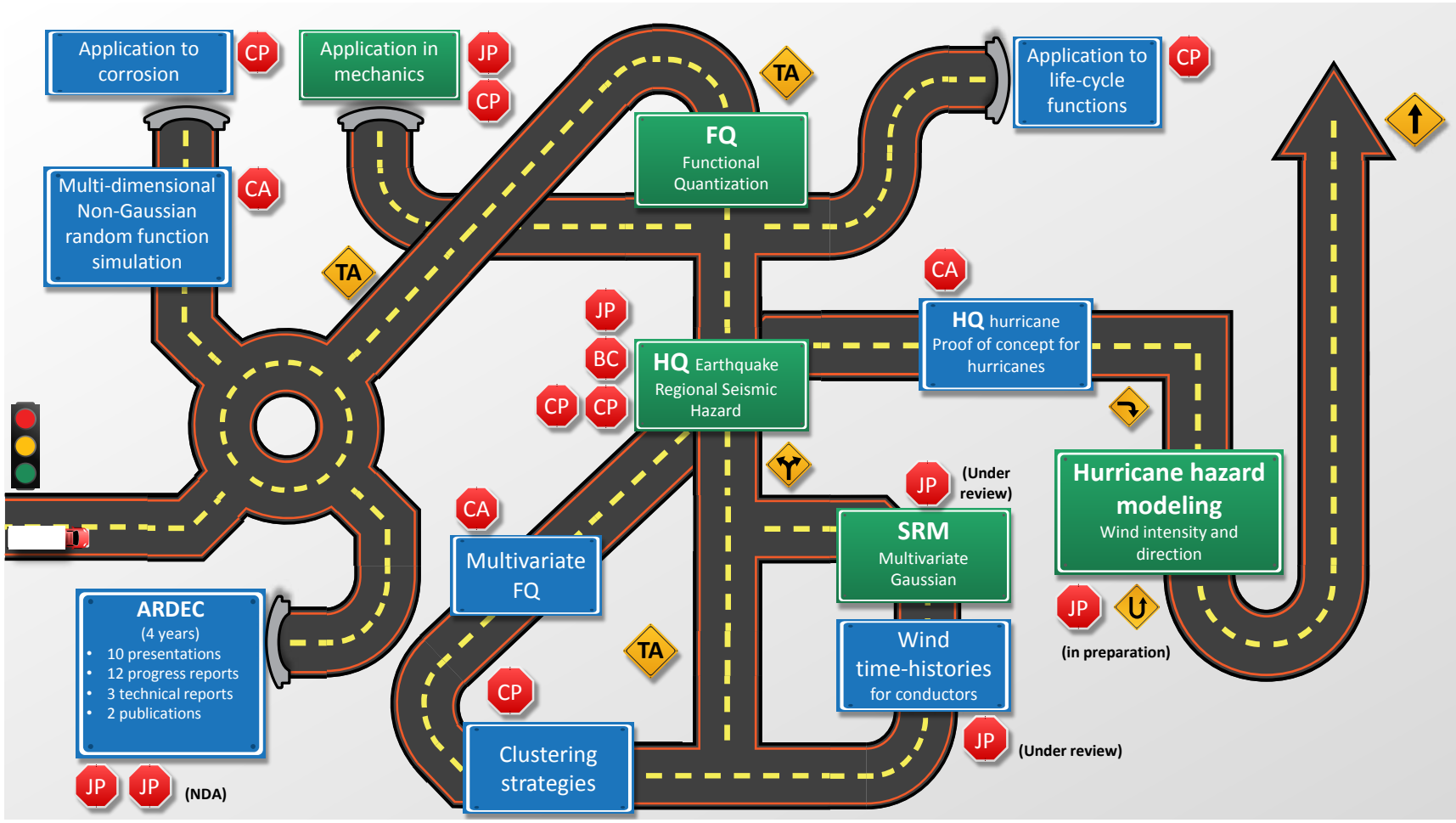


Figure 1.2: Road map during the Ph.D. program. The blue signs indicate research work not explicitly discussed in the following chapters. The green signs indicate the research presented in this thesis. The red "JP", "CP", "BC", "CA" stop signs indicate a journal article, conference paper, book chapter and conference abstract, respectively.

- Constructed fault tree models and established reliability block diagrams
- Filtered experimental data and fitted parametric and non-parametric probability density functions to them
- Developed a system reliability model of the weapon's mechanical system using several performance indicators, such as the probability mass function (PMF) and the cumulative distribution function (CDF) of the system time-to-failure, the survivor and hazard functions and the importance ranking of failure modes
- Developed a dynamic model which considers the effect of component maintenance actions to the lifetime behavior of a weapon system
- Implemented a generalized (Gaussian and non-Gaussian) correlated random number simulator to investigate the effect of correlation between components' life-time-functions.
- Performed numerical simulation to asses the accuracy of the developed methodology and cross-validated it with experimental data

In addition to the tasks above, throughout the project, the researcher contributed to the preparation and delivery of 10 presentations, the submission of 12 progress reports, the submission of 3 technical reports and the submission of 2 manuscript for potential journal publication, which are still covered by a non-disclosure agreement (NDA) (Frangopol et al., 2015b,a).

One of the first research tasks in the simulation of random fields and processes conducted by the researcher was the development of a methodology for the **simulation of multi-dimensional non-Gaussian random fields**. The technique is an extension to the case of multi-dimensional random functions of a recently developed iterative technique that generates one-dimensional, univariate sample functions of non-Gaussian random fields and processes (Shields et al., 2011). A conference paper was published, in which the methodology was presented and used to model probabilistically the spatial distribution of corrosion

of a steel beam (Christou and Bocchini, 2014a). The amount of corrosion penetration was modeled as a random field and applied to a finite element mesh on the top flange of a steel cantilever beam. The field matches both the arbitrarily prescribed spectral density function and the non-Gaussian marginal distribution. The effect of the spatial variability of the corrosion on the flange was evaluated using the tip displacement as representative structure response metric, while a series of finite element analysis were performed by a Monte Carlo simulation. The flanges and web were modeled with 4-node Serendipity shells that assume bilinear deflection, bilinear rotations and linear transverse shear strain. To effectively and efficiently consider the effect of field variability at the finite element thickness reduction, *an in-house three-dimensional finite element software was developed by the author in MATLAB* (MATLAB, 2010) which was able to consider different structures made of plates and shells elements.

After familiarizing himself with random processes via the methodology above, the researcher then began to develop a technique to sample random functions more effectively than traditional Monte Carlo simulations (MCS). The goal of the technique, called "**Functional Quantization (FQ)**" is to optimally represent random fields and processes using a finite number of samples (Luschgy and Pagès, 2002). The technique has proven to optimally approximate random functions using a pre-determined number N of representative samples. Functional Quantization is distinguishable from MCS because of the following: (1) the representative samples from FQ are selected not entirely at random and (2) the representative samples from FQ are not equally weighted. A limited number of techniques for the selection of optimal samples and computation of associated probabilities based on the FQ concept have been presented in the literature; however, they have only been demonstrated on Gaussian random functions. The researcher investigated a recently developed technique called "Functional Quantization by Infinite-Dimensional Centroidal Voronoi Tessellation" (FQ-IDCVT) which has been successfully used for one-dimensional,

non-Gaussian and non-stationary processes (Miranda and Bocchini, 2015a) and extended the technique to the case of random functions defined over a multi-dimensional domain (Christou et al., 2016b; Christou and Bocchini, 2014b). The description of the modifications to the existing algorithm required by the extension to multi-dimensional fields is accompanied by its demonstration on two numerical applications. The first example involves a two-dimensional lognormal field, generated through the Spectral Representation Method (SRM) using the previously developed iterative algorithm, which approximates a non-Gaussian stationary multi-dimensional field, to investigate the limits of applicability and discuss the computational challenges associated with the extension to two-dimensional fields. The second numerical application was in the field of computational mechanics and involved a two-dimensional panel in plane-stress with uncertain Young modulus, modeled as a two-dimensional stochastic field.

Having investigated the limits of applicability of the FQ-IDCVT technique for univariate and multivariate random processes and fields, it was expected that it would yield promising results when applied to problems involving random functions with relatively high correlation length. Therefore, its application to **quantize probabilistic life-cycle performance models** was conducted. In life-cycle engineering, the uncertainty in the deteriorating performance (e.g., reliability) of structures is often described by means of time-dependent models with random parameters. Most of the numerical models of individual structures are already complex, and when a life-cycle performance analysis of a network of systems is carried out, the complexity and the computational cost become very high. Thus, for simulation-based probabilistic analysis, the number of deterministic runs that can actually be performed is limited. The FQ-IDCVT technique was applied to the optimal selection of life-cycle profile samples, which are non-Gaussian and non-stationary random functions (Bocchini et al., 2014). The results were assessed and compared with those from MCS and Latin Hypercube Sampling (LHS) techniques. In particular, LHS has

been proven to be a very effective technique to sample this type of models, because they are functions of a small number of parameters. However, the accuracy of the representation of the life-cycle models is significantly enhanced when FQ-IDCVT is used, compared to MCS and even LHS, thus making this simulation approach applicable even when the number of deterministic runs needs to be small.

Extending the FQ-IDCVT technique to the case of random functions defined over two dimensional domains was the stepping stone to develop a methodology for the selection of an optimal set of stochastic intensity measure (IM) maps representing the regional hazard over a geographic area. As mentioned, the purpose of selecting a representative set of IM maps is to use them for the analysis of spatially distributed infrastructure systems. An approach was proposed and called "**Hazard Quantization (HQ)**" (Christou and Bocchini, 2015; Bocchini et al., 2016; Karamlou et al., 2016). HQ embraces the nature of regional IM maps as two-dimensional random fields and ensures mean-square convergence of the ensemble of representative IM maps to the complete portfolio of possible hazard events, which is a particularly important property for risk analysis. The technique's applicability was demonstrated for the regional seismic hazard analysis of the Charleston, South Carolina region (Christou et al., 2017). A small set of IM maps and their associated probabilities resulting for the application of HQ are evaluated at all points and all pairs of points, on their ability to correctly represent the hazard curve and the IM autocorrelation. In addition, a detailed comparison of the proposed technique with other popular methodologies in the same field is presented, showing that HQ in general provides results comparable or superior to the best techniques available.

Applying the FQ-IDCVT technique to regional hazard analysis problems demonstrated that the technique works effectively well; however, the computational cost often becomes an issue when multi-dimensional fields are at hand. This is because the computational time scales linearly with the size of the domain and the resolution that is used. The increase in the

computational cost results from the distance calculations between the samples. To address the issue of the escalating computational cost, the implementation of several **accelerated clustering techniques** was investigated (Fiorillo et al., 2017). The results of each algorithm were used to perform an accuracy and efficiency comparison of each methodology against a benchmark solution.

The following research task was to extend the HQ methodology to the multivariate case. This would require a **multivariate random function generator**. Therefore, a well-known version of the Spectral Representation Method (SRM) was investigated for the simulation of multivariate random processes and fields (Deodatis, 1996). During the implementation of the proposed algorithm, it was realized that the Cross-Spectral Density Matrix (CSDM) of the generated samples does not match the values of the CSDM at the discrete frequencies which are considered as input in the simulation procedure (, Christou and Bocchini). This observation was investigated mathematically and yielded a closed form expression which proves that even in the case of samples with finite period (or discrete frequency domain), these samples are ergodic in the correlation. In addition, it was also showcased that in the limit, the derived closed-form expression of the sample CSDM coincides with the input CSDM.

The findings on the SRM algorithm assisted in the accurate simulation of multivariate hurricane wind speeds. This process captures the uncertainties of the hurricane wind fluctuations and correlation in time and space for a significantly large number of components (i.e., random load time histories at specific locations). The hurricane wind fluctuations were used as input for deriving the fragility model of electrical conductors (, Ma et al.).

The following research task was to **apply FQ-IDCVT to vector processes describing wind fluctuations**, corresponding to different locations over the façade of a building or along the length of a conductor power line (Christou et al., 2016a).

Keeping in mind that an important goal is to develop fundamental hazard models for multi-hazard risk assessment, the **HQ methodology was extended and applied to the case of regional hurricane hazard analysis** of a region in the southeastern United States (Christou and Bocchini, 2018). The extension of the HQ technique included the ability to preselect a fixed number of real events to include in the regional analysis a priori and then perform the hazard quantization for the optimal selection of the remaining set of hurricane events.

To apply the hazard quantization methodology, a high number of available samples is required, such that they can truly represent the sample space. In the case of hurricane hazard, the number of historical events is not enough to effectively apply HQ. A collection of synthetic events was kindly shared by Professor Kerry Emanuel (MIT), and it was used for a proof of concept, but it was still insufficient for thorough testing of the methodology. This significant constraint led the researcher to the **development of an in-house simulation framework of synthetic hurricane directional wind speeds**. This framework is presented in the following chapters in detail and can be used to feed the HQ technique with the required number of scenario events and further used for hazard and risk analysis in the United States.

1.3 Research objectives

This research proposes new methodologies to probabilistically represent stochastic processes, random fields, and natural hazards. These methodologies have been demonstrated in applications aimed at reducing risk and the impact of natural disasters on vulnerable communities.

The main objectives of this work are to:

- develop accurate simulators of regional intensity measures for earthquakes (on the basis of seismological models and historical data) and hurricanes (genesis, track propagation, storm intensity, key atmospheric parameters, wind field, and dissipation, on the basis of physics-based and data-driven approaches);
- propose a technique for the optimal selection of a limited number of intensity measure maps to use as input in a regional risk and resilience analysis of interdependent infrastructure systems;
- introduce more comprehensive ways to represent regional hazards, such as the combination of wind direction and intensity for hurricanes;
- assess the accuracy of the proposed methodologies, validate them against the available data, and compare them with the state-of-the-art techniques.

1.4 Outline of the Dissertation

The dissertation comprises six chapters. Each chapter's content is discussed below.

Chapter 1, the current chapter, details the research overview and the research objectives.

Chapter 2 investigates an algorithm for the simulation of multivariate turbulent hurricane wind velocity fluctuations. This simulation can be used to model hurricane-induced dynamic loads on power transmission networks. Included in this chapter is a review of a commonly used in practice algorithm for the simulation of multivariate Gaussian stationary random processes. The researcher demonstrated a closed-form expression indicating that when the frequency domain is discretized the generated samples are ergodic; however, the cross-spectral density matrix output does not match the input CSDM. Finally, the closed-form expression of the generated samples' resulting spectrum is provided, along with a

delineation of surprising expression characteristics, such as the fact that the variance of the generated samples is influenced by the algorithm's frequency indexing technique.

Chapter 3 presents techniques for the representation of random fields with a finite number of samples and their application to mechanics. A methodology is proposed for the optimal selection of a moderate number of samples effectively representing the entire space of sample realizations. The technique is then applied to multidimensional random functions, such as two-dimensional non-Gaussian random fields and a two-dimensional panel with uncertain Young modulus under plane stress.

Chapter 4 investigates the seismic characterization of the Charleston, South Carolina region and presents a model for the simulation of the regional seismic hazard. Additionally, a methodology —named Hazard Quantization (HQ)— is proposed for the selection of an optimal set of stochastic simulated and/or historical seismic intensity measure (IM) maps representing the regional hazard over a geographic area, and applied to the city of Charleston. Two state-of-the-art methodologies are implemented and compared as a means of assessing the strengths and weaknesses of the proposed methodology. Lastly, a quantitative comparison utilizing various metrics as well as a qualitative comparison of framework features is discussed.

Chapter 5 proposes a methodology to implement a hurricane wind simulator for the assessment of directional hurricane risk over the Atlantic and Gulf coasts of the United States. Implementation details are presented and probabilistic modeling module are discussed. Such modules include: the hurricane track, the hurricane intensity, the hurricane intensity decay, the radius to maximum wind, Holland B parameter, the wind field model, the boundary layer model and the wind directionality. The chapter discusses the available models for each of the aforementioned modules and presents the models used in this framework. When feasible, each module is assessed to investigate the limits of applicability and

the epistemic uncertainty. The historical data used are presented along with the required operations needed to prepare them as input for the required probabilistic models.

Chapter 6 presents the summary of the dissertation and the significant contributions of the conducted research.

Chapter 2

On the Spectrum of Multi-Variate Gaussian, Ergodic, Random Samples Generated by Spectral Representation

2.1 Introductory remarks

A large amount of problems in stochastic mechanics and engineering involve random processes and fields as for example the wind velocity fluctuations along a transmission power line. Often, simulation-based approaches are used to solve such problems. Therefore, effective techniques for the accurate and efficient generation of random samples matching the prescribed characteristics are needed. Among the various techniques available in the literature, the Spectral Representation Method (SRM) is one of the most popular in science and practice (Shinozuka and Jan, 1972). One of the most interesting features of the SRM is that it can generate samples that are ergodic in the limit, when the spectrum is modeled over a continuous frequency domain (Shinozuka and Deodatis, 1991).

Deodatis (1996) proposed an extension of the classical SRM that generates multi-variate samples. In the same paper, the author suggested to use the Frequency Double-Indexing (FDI) technique proposed by Zerva (1992), and he showed that in the limit, when the samples have infinite period and the spectra are modeled over a continues frequency domain, the generated samples are ergodic. However, in practice, these simulations are always performed through numerical evaluations, so the spectra are defined at discrete frequencies separated by finite intervals. The first original contribution of this chapter is the proof that the CSDM of the generated samples (named here "*actual* CSDM") does not

match the values of the CSDM at the discrete frequencies which are considered as input in the simulation procedure, named for the remainder of the chapter as "*input* CSDM". Therefore, interesting properties of the generated samples, such as their variance, are influenced by the simulation algorithm and specifically by the FDI technique.

The second original contribution of this chapter is the proof is the generalization of the proof of ergodicity in the correlation, , to show that it applies not only in the (theoretical) limit, but also in practical cases where the frequency domain has discrete frequencies and the samples have finite period. This means that all samples have exactly the same Cross-Spectral Density Matrix (CSDM), and therefore the CSDM can be computed by space-averaging (or temporal-averaging) over one sample, rather than ensemble-averaging multiple samples.

The third original contribution of this chapter is that itprovides a closed-form expression that allows to compute a priori the *actual* CSDM of the generated samples, which permits the analyst to compute ahead of the simulation the approxiation error on the CSDM.

In the reminder of the chapter, the case of random fields defined over the space domain is considered. The extension to random processes defined over the time domain is straightforward.

2.2 Revisiting the multi-variate SRM

The objective of the multi-variate SRM is to generate Gaussian samples that match a prescribed stationary *theoretical* CSDM, $\mathbf{S}(\kappa)$, defined over a continuous wave number do-

main κ:

$$\mathbf{S}^0(\kappa) = \begin{bmatrix} S_{11}(\kappa) & S_{12}(\kappa) & \cdots & S_{1m}(\kappa) \\ S_{21}(\kappa) & S_{22}(\kappa) & \cdots & S_{2m}(\kappa) \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1}(\kappa) & S_{m2}(\kappa) & \cdots & S_{mm}(\kappa) \end{bmatrix} \tag{2.1}$$

where κ represents the wave number (i.e., the frequency domain, when dealing with random fields defined over space), and m is the total number of variates. In the following, a unidimensional (1D), m-variate (mV) random vector field $\boldsymbol{g}(\kappa)$ with components $g_j(x);\ j = 1,\ldots,m$ having mean value equal to zero

$$E\left[g_j(x)\right] = 0; \quad j = 1,\ldots,m$$

and a prescribed (*target*) cross-spectral density matrix given by Eq. (2.1), is considered. Distinction is made between the vector field $\boldsymbol{g}(x)$ and its sample generation $\hat{\boldsymbol{g}}(x)$.

2.2.1 Decomposition of the target CSDM

In order to simulate a 1D-mV random field using SRM, its CSDM must first be decomposed into the following product:

$$\mathbf{S}(\kappa) = \mathbf{H}(\kappa)\mathbf{H}^{T*}(\kappa) \tag{2.2}$$

where * indicates the complex conjugate and superscript T the transpose. Such decomposition can be performed in several ways, for instance using the Cholesky method as suggested by Deodatis Deodatis (1996). When Chelosky's decomposition is used, the terms of matrix

$\mathbf{H}$ can be computed recursively as follows:

$$\begin{aligned} H_{cc}(\kappa) &= \sqrt{S_{cc}(\kappa) - \sum_{e=1}^{c-1} H_{ce}(\kappa) H_{ce}^{*}(\kappa)} \\ H_{cd}(\kappa) &= \frac{1}{H_{dd}(\kappa)} \left(S_{cd}(\kappa) - \sum_{e=1}^{d-1} H_{ce}(\kappa) H_{de}^{*}(\kappa) \right), \quad \text{for} \quad c > d \end{aligned} \tag{2.3}$$

In this decomposition $\mathbf{H}(\kappa)$ is a lower triangular matrix:

$$\mathbf{H}(\kappa) = \begin{bmatrix} H_{11}(\kappa) & 0 & \cdots & 0 \\ H_{21}(\kappa) & H_{22}(\kappa) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H_{m1}(\kappa) & H_{m2}(\kappa) & \cdots & H_{mm}(\kappa) \end{bmatrix} \tag{2.4}$$

whose diagonal elements are real and non-negative functions of κ and whose off-diagonal elements are complex functions of κ, in general. The off-diagonal elements $H_{cd}(\kappa)$ can also be written in polar form as:

$$H_{cd}(\kappa) = |H_{cd}(\kappa)| \; e^{i\theta_{cd}(\kappa)}$$

where

$$\theta_{cd}(\kappa) = \tan^{-1} \left\{ \frac{\mathrm{Im}\left[H_{cd}(\kappa)\right]}{\mathrm{Re}\left[H_{cd}(\kappa)\right]} \right\} \tag{2.5}$$

2.2.2 Generation of random samples

Once matrix $\mathbf{H}(\kappa)$ has been assessed, samples can be generated using the following series of steps, which are a slightly revised version of what presented in Deodatis (1996). First,

vectors $t_{cd}^{(i)}$ are generated as:

$$t_{cd}^{(i)}(p\,\Delta x) = \mathrm{DFT}^{+}\left[2H_{cd}\left(n\Delta\kappa + \frac{d}{m}\Delta\kappa\right)\sqrt{\Delta\kappa}\cdot\exp\left(\mathrm{i}\phi_{dn}^{(i)}\right)\right]; \qquad n = 0,1,\ldots,M-1 \tag{2.6}$$

where the indexes run as follows

$$c = 1,2,\ldots,m; \quad d = 1,2,\ldots,c; \quad p = 0,1,\ldots,M-1;$$

and where Δx defines the discretization in space. Indexes c and d are used for the remainder of the chapter as generic indices to identify a component and the component's term respectively of the random field. $\mathrm{DFT}^{+}[\cdot]$ is the Discrete Fourier Transform defined as:

$$\mathrm{DFT}^{+}[\mathrm{Y}] = \sum_{\alpha=1}^{Z} \mathrm{Y}(\kappa_\alpha)\, e^{\frac{2\pi}{Z}\mathrm{i}(\beta-1)(\alpha-1)} = \mathrm{X}(x_\beta); \qquad \beta = 1,\ldots,Z \tag{2.7}$$

and $\mathrm{DFT}^{-}[\cdot]$, as it will be used later, is defined as:

$$\mathrm{DFT}^{-}[\mathrm{X}] = \sum_{\beta=1}^{Z} \mathrm{X}(x_\beta)\, e^{-\frac{2\pi}{Z}\mathrm{i}(\beta-1)(\alpha-1)} = \mathrm{Y}(\kappa_\alpha); \qquad \alpha = 1,\ldots,Z \tag{2.8}$$

X and Y are vectors of length Z and i indicates the imaginary unit. Parameter $\phi_{dn}^{(i)}$ defines the random phase angles which are uniformly distributed in the interval $[0, 2\pi]$ and superscript (i) refers to the i-th set of phase angle samples and indicates that a specific set of random phase angles yields a specific sample $t_{cd}^{(i)}$. It should be noted that Eq. (2.6) assumes that both the space domain and the wave number domain are discretized, with steps Δx and $\Delta\kappa$ respectively, which is always the case when numerical analysis is used. The step in the wave number domain, $\Delta\kappa$, is defined as follows:

$$\Delta\kappa = \frac{\kappa_u}{N} \tag{2.9}$$

where κ_u is the upper cut-off wave number and N is the number of discretization points in the wave number domain. The upper cut-off wave number needs to be selected such that the values of the elements of the CSDM beyond κ_u may assumed to be zero (i.e., $S_{cd}(r\Delta\kappa) = 0$ for $r\Delta\kappa > \kappa_u = N\Delta\kappa$). The step increment, Δx, in the space domain is computed as:

$$\Delta x = \frac{2\pi}{M\Delta\kappa} \tag{2.10}$$

so that aliasing is avoided.

Next, m identical copies of $t_{cd}^{(i)}$ are concatenated to generate $h_{cd}^{(i)}$:

$$h_{cd}^{(i)}(q\Delta x) = t_{cd}^{(i)}(\mathrm{mod}(q,M)\Delta x); \quad q = 0,1,...,mM-1 \tag{2.11}$$

where $\mathrm{mod}(q,M)$ denotes the remainder of division of q by M.

Because of the frequency double indexing Deodatis (1996), $h_{cd}^{(i)}$ has to be multiplied by a shifting factor:

$$\begin{aligned} \hat{h}_{cd}^{(i)}(q\Delta x) &= \Re\left\{ h_{cd}^{(i)}(q\Delta x) \cdot \exp\left[\mathrm{i}\frac{d}{m}\Delta\kappa\,(q\Delta x) \right] \right\} \\ &= \Re\left\{ h_{cd}^{(i)}(q\Delta x) \cdot \exp\left(\mathrm{i}\frac{2\pi}{M}\, q\frac{d}{m} \right) \right\} \end{aligned} \tag{2.12}$$

where $\Re\{\cdot\}$ denotes the real part of the quantity in brackets.

Finally, each component c of sample (i) of the random field is obtained as the superposition of functions $\hat{h}_{cd}^{(i)}(q\Delta x)$:

$$\hat{g}_c^{(i)}(q\Delta x) = \sum_{d=1}^{c} \hat{h}_{cd}^{(i)}(q\Delta x); \quad c = 1,2,...,m \tag{2.13}$$

Sample $\hat{\boldsymbol{g}}^{(i)}(q\Delta x)$ and its m components are the final result of the simulation process.

2.3 Cross-Spectra Density Matrix estimation of the generated samples

When a sample $\hat{\boldsymbol{g}}^{(i)}(q\Delta x)$ is generated as prescribed in Section 2.2, the Cross-Spectral Density Function (CSDF) of any pair of components $\hat{g}_r^{(i)}(q\Delta x)$ and $\hat{g}_s^{(i)}(q\Delta x)$ of the sample can be estimated in the temporal sense at discrete frequencies $l\frac{\Delta\kappa}{m}$, where $l = 0,1,2,\ldots,mM-1$ using the formula (Bendat and Piersol, 1986):

$$S_{rs}^{g}\left(l\frac{\Delta\kappa}{m}\right) = \frac{1}{2\pi(mL)}\mathrm{DFT}^{-}[\hat{g}_r^{(i)}(q\Delta x)]\,\Delta x\left\{\mathrm{DFT}^{-}[\hat{g}_s^{(i)}(q\Delta x)]\,\Delta x\right\}^{*} \tag{2.14}$$

where $\mathrm{DFT}^{-}[\cdot]$ denotes the Discrete Fourier Transform with negative exponent as defined by Eq. (2.8), the asterisk denotes the complex conjugate, and $L = \frac{2\pi}{\Delta\kappa}$ is the length of the sample. Indices r and s are generic indices and represent two of the components of the vector field. When the DFT is performed, the result has the same number of points as the input function $\hat{\boldsymbol{g}}(q\Delta x)$; in this case mM points. Substituting Eq. (2.13) into Eq. (2.14) and taking advantage of the linear property of the DFT operator, we get:

$$\begin{aligned} S_{rs}^{g}\left(l\frac{\Delta\kappa}{m}\right) &= \frac{\Delta x^2}{2\pi(mL)}\mathrm{DFT}^{-}\left[\sum_{j=1}^{r}\hat{h}_{rj}^{(i)}(q\Delta x)\right]\left\{\mathrm{DFT}^{-}\left[\sum_{k=1}^{s}\hat{h}_{sk}^{(i)}(q\Delta x)\right]\right\}^{*} \\ &= \frac{\Delta x^2}{2\pi(mL)}\sum_{j=1}^{r}\left[\mathrm{DFT}^{-}\left[\hat{h}_{rj}^{(i)}(q\Delta x)\right]\right]\left\{\sum_{k=1}^{s}\left[\mathrm{DFT}^{-}\left[\hat{h}_{sk}^{(i)}(q\Delta x)\right]\right]\right\}^{*} \end{aligned} \tag{2.15}$$

Equation. (2.15) is simplified by first expanding function $\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right]$, where c can refer to any component of the vector field and d refers to one of the component's term. Using Eq. (2.8) and recalling that $\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right]$ is a function of l with mM points,

then $\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right]$ can be written in:

$$\begin{aligned}\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right] &= \sum_{q=0}^{mM-1} \hat{h}_{cd}^{(i)}(q\Delta x)\, e^{-\mathrm{i}\frac{2\pi}{mM}\,q\,l} \\ &= \sum_{v=1}^{m} \sum_{q=(v-1)M}^{vM-1} \mathrm{Re}\left\{ h_{cd}^{(i)}(q\Delta x)\, e^{\mathrm{i}\frac{2\pi}{mM}\,q\,d} \right\} e^{-\mathrm{i}\frac{2\pi}{mM}\,q\,l}\end{aligned} \tag{2.16}$$

Changing the variable q with $p = q-(v-1)M$, Eq. (2.16) becomes:

$$\begin{aligned}\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right] = \sum_{v=1}^{m}\sum_{p=0}^{M-1} \mathrm{Re}\left\{ t_{cd}^{(i)}(p\Delta x) \exp\left[\mathrm{i}\frac{2\pi}{M}\,p\,\frac{d}{m}\right] \exp\left[\mathrm{i}\,2\pi\frac{d}{m}\,(v-1)\right]\right\} \\ \cdot \exp\left[-\mathrm{i}\frac{2\pi}{M}\,p\,\frac{l}{m}\right] \exp\left[-\mathrm{i}\,2\pi\,\frac{l}{m}\,(v-1)\right]\end{aligned} \tag{2.17}$$

When defining $\tilde{t}_{cd}^{(i)}(p\Delta x) = t_{cd}^{(i)}(p\Delta x)\exp\left[\mathrm{i}\frac{2\pi}{M}\,p\,\frac{d}{m}\right]$ and extracting the real part of the complex expression in the braces using its conjugate, Eq. (2.17) can be written as follows:

$$\begin{aligned}\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right] = \underbrace{\frac{1}{2}\sum_{p=0}^{M-1} \tilde{t}_{cd}^{(i)}(p\Delta x)\, e^{-\mathrm{i}\frac{2\pi}{M}\,p\,\frac{l}{m}}}_{A} \underbrace{\sum_{v=0}^{m-1} e^{-\mathrm{i}\,2\pi\,v\left(\frac{l}{m}-\frac{d}{m}\right)}}_{B} \\ + \frac{1}{2}\sum_{p=0}^{M-1} \left[\tilde{t}_{cd}^{(i)}(p\Delta x)\right]^{*} e^{-\mathrm{i}\frac{2\pi}{M}\,p\,\frac{l}{m}} \underbrace{\sum_{v=0}^{m-1} e^{-\mathrm{i}\,2\pi\,v\left(\frac{l}{m}+\frac{d}{m}\right)}}_{C}\end{aligned} \tag{2.18}$$

Expression B of Eq. (2.18) can be further analysed considering the Euler's formula:

$$\begin{aligned}B &= \sum_{v=0}^{m-1}\left[\cos\left(2\pi\,v\,\frac{l-d}{m}\right) - \mathrm{i}\sin\left(2\pi\,v\,\frac{l-d}{m}\right)\right] \\ &= \begin{cases} m & l = km+d; \quad k = 0,1,\ldots,M-1 \\ 0 & \text{otherwise}\end{cases}\end{aligned} \tag{2.19}$$

The summation over the cosines in Eq. (2.19) is computed as:

$$\sum_{v=0}^{m-1} \cos\left(2\pi v \frac{l-d}{m}\right) = \begin{cases} m & l = km+d; \quad k = 0,1,\ldots,M-1 \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

The summation over the sines in Eq. (2.19) is computed as:

$$\sum_{v=0}^{m-1} \sin\left(2\pi v \frac{l-d}{m}\right) = 0 \quad \forall\, l,d \tag{2.21}$$

Expression A of Eq. (2.18) can be simplified using Eq. (2.6) as follows:

$$\begin{aligned} A &= \sum_{p=0}^{M-1}\sum_{n=0}^{M-1} 2H_{cd}^{*}\left[\left(n+\frac{d}{m}\right)\Delta\kappa\right]\sqrt{\Delta\kappa}\, e^{\mathrm{i}\phi_{d,n}^{(i)}} e^{\mathrm{i}\frac{2\pi}{M}p\left(n+\frac{d}{m}-\frac{l}{m}\right)} \\ &= \sum_{n=0}^{M-1} 2H_{cd}^{*}\left[\left(n+\frac{d}{m}\right)\Delta\kappa\right]\sqrt{\Delta\kappa}\, e^{\mathrm{i}\phi_{d,n}^{(i)}} \underbrace{\sum_{p=0}^{M-1} e^{\mathrm{i}\frac{2\pi}{M}p\left(n+\frac{d}{m}-\frac{l}{m}\right)}}_{D} \\ &= 2MH_{cd}^{*}\left(l\frac{\Delta\kappa}{m}\right)\sqrt{\Delta\kappa}\, e^{\mathrm{i}\phi_{d,\frac{l-d}{m}}^{(i)}} \end{aligned} \tag{2.22}$$

Recalling from Eq. (2.19) that parameter l can only be of the form $l = km+d; k = 0,1,\ldots,M-1$ in order for the product $A \cdot B$ of Eq. (2.18) to be different than zero. Therefore, parameter l should satisfy $l = nm+d;\ n = 0,1,\ldots,M-1$ and for the pair of l,n that this holds, expression D becomes: $\sum_{p=0}^{M-1} e^{-\frac{2\pi}{M}p\left(n+\frac{d}{m}-\frac{l}{m}\right)} = 1$. For all pairs of n,l which satisfy $l = nm+d;\ n = 0,1,\ldots,M-1$, expression A yields the last equality of Eq. (2.22).

Expression C in Eq. (2.18) is zero for all values of l and d.

$$\begin{aligned} C &= \sum_{v=0}^{m-1}\left[\cos\left(2\pi v\frac{l+d}{m}\right) - \mathrm{i}\sin\left(2\pi v\frac{l+d}{m}\right)\right] \\ &= 0 \quad \forall\, l,d \end{aligned} \tag{2.23}$$

Both the real and imaginary parts in the equation above are equal to zero for all values of l

and d. Therefore, making use of Eqs. (2.19)-(2.23), Eq. (2.18) becomes:

$$\mathrm{DFT}^{-}\left[\hat{h}_{cd}^{(i)}(q\Delta x)\right] = mMH_{cd}^{*}\left(l\frac{\Delta\kappa}{m}\right)\sqrt{\Delta\kappa}\,e^{\mathrm{i}\phi_{d,\frac{l-d}{m}}^{(i)}} \tag{2.24}$$

Considering Eqs. (2.15) and (2.24) we can observe that the product of the sums gives terms different from zero only when both factors $\mathrm{DFT}^{-}\left[\hat{h}_{rj}\right]$ and $\left\{\mathrm{DFT}^{-}\left[\hat{h}_{sj}\right]\right\}^{*}$ are different from zero. For this reason is possible to consider only terms with $j=k$ and, therefore, simplify one summation of Eq. (2.15). Equation (2.15) now becomes:

$$S_{rs}^{g}\left(l\frac{\Delta\kappa}{m}\right) = \sum_{j=1}^{\min(r,s)}\frac{\Delta x^2}{2\pi(mL)}\left[\mathrm{DFT}^{-}\left[\hat{h}_{rj}^{(i)}(q\Delta x)\right]\right]\left[\mathrm{DFT}^{-}\left[\hat{h}_{sj}^{(i)}(q\Delta x)\right]\right]^{*} \tag{2.25}$$

Now, for any j we can compute the corresponding term in the summation in Eq. (2.25) as follows:

$$\begin{aligned}&\frac{\Delta x^2}{2\pi(mL)}\mathrm{DFT}^{-}\left[\hat{h}_{rj}^{(i)}(q\Delta x)\right]\left\{\mathrm{DFT}^{-}\left[\hat{h}_{sj}^{(i)}(q\Delta x)\right]\right\}^{*} = \\ &= \begin{cases} mH_{rj}^{*}\left(l\frac{\Delta\kappa}{m}\right)H_{sj}\left(l\frac{\Delta\kappa}{m}\right) & l = nm+j;\ n=0,1,\ldots,M-1 \\ 0 & \text{otherwise}\end{cases}\end{aligned} \tag{2.26}$$

It should be noted here that the value at zero frequency (i.e., $l=0$) will always result to zero as expected. Finally, Eq. (2.15) becomes:

$$S_{rs}^{g}\left(l\frac{\Delta\kappa}{m}\right) = \begin{cases} mH_{r1}^{*}\left(l\frac{\Delta\kappa}{m}\right)H_{s1}\left(l\frac{\Delta\kappa}{m}\right) & l = nm+1 \\ mH_{r2}^{*}\left(l\frac{\Delta\kappa}{m}\right)H_{s2}\left(l\frac{\Delta\kappa}{m}\right) & l = nm+2 \\ \quad\vdots & \\ mH_{r\min(r,s)}^{*}\left(l\frac{\Delta\kappa}{m}\right)H_{s\min(r,s)}\left(l\frac{\Delta\kappa}{m}\right) & l = nm+\min(r,s) \\ 0 & \text{otherwise}\end{cases} \tag{2.27}$$

where $n = 0, 1, \ldots, M-1$.

Because of the discretization in the frequency domain and the FDI, S_{rs}^{g} can have energy at all or some of the frequencies in $l\frac{\Delta\kappa}{m}$. In particular, at frequencies $n\Delta\kappa + \frac{j}{m}\Delta\kappa$, the spectrum S_{rs}^{g} is a delta function, while at the other frequencies it is zero. Therefore, Eq. (2.27) can be written more condensely:

$$\begin{aligned} S_{rs}^{g}\left(l\frac{\Delta\kappa}{m}\right) &= mH_{rj}^{*}\left(l\frac{\Delta\kappa}{m}\right)H_{sj}\left(l\frac{\Delta\kappa}{m}\right)\cdot\delta\left[l\frac{\Delta\kappa}{m}-\left(n\Delta\kappa+\frac{j}{m}\Delta\kappa\right)\right] \\ &\qquad j=1,2,\ldots,\min(r,s); \quad n=0,1,\ldots,M-1 \\ &= mH_{rj}^{*}\left(l\frac{\Delta\kappa}{m}\right)H_{sj}\left(l\frac{\Delta\kappa}{m}\right)\cdot\delta\left[l-nm-j\right] \\ &\qquad j=1,2,\ldots,\min(r,s); \quad n=0,1,\ldots,M-1 \end{aligned} \tag{2.28}$$

where $l = 0, 1, 2, \ldots, mM-1$. Equation (2.28) is a closed-form expression of the CSDM that can be obtained when computing the empirical spectra from a single sample, starting from Eq. (2.14) and making use of the DFT technique. The single sample is also considered to have been generated using the SRM and the DFT technique as presented in Sec. 2.2.2. It should be mentioned that when the DFT is performed for the CSDM estimation, the result has as many points as the input function $\hat{\boldsymbol{g}}(q\Delta x)$; however, only the first half of the resulting points is meaningful, while the second half is only its mirror image due to the periodicity implicit in the use of DFT. Considering now mN points and letting $\bar{\boldsymbol{\kappa}} = [0\ 1\ 2\ \ldots\ mN-1]^{T}\cdot\Delta\kappa$, Eq. (2.28) can also be expressed as:

$$\begin{aligned} S_{rs}^{g}(\bar{\boldsymbol{\kappa}}) &= mH_{rj}^{*}(\bar{\boldsymbol{\kappa}})H_{sj}(\bar{\boldsymbol{\kappa}})\cdot\delta\left[\bar{\boldsymbol{\kappa}}-\left(n\Delta\kappa+\frac{j}{m}\Delta\kappa\right)\right] \\ &\qquad j=1,2,\ldots,\min(r,s); \quad n=0,1,\ldots,N-1 \end{aligned} \tag{2.29}$$

Equations (2.28) and (2.29) compute the spectrum of the samples as it is naturally given by their digital representation, that means at intervals $\frac{\Delta\kappa}{m}$. To measure the spectral density over intervals of width $\Delta\kappa$ (i.e., at the primary discretization wave numbers), we can average the

results at m sub-frequencies (i.e., sum the areas under the SDF). Hence, Eq. (2.29) can be expressed in terms of the new domain $\tilde{\boldsymbol{\kappa}} = [0\ 1\ 2\ \ldots\ N]^T \cdot \Delta\kappa$ and thus becomes:

$$S_{rs}^{g}(\tilde{\boldsymbol{\kappa}}) = H_{rj}^{*}(\bar{\boldsymbol{\kappa}})\, H_{sj}(\bar{\boldsymbol{\kappa}}) \cdot \delta\left[\bar{\boldsymbol{\kappa}} - \left(n\Delta\kappa + \frac{j}{m}\Delta\kappa\right)\right] \cdot \delta\left[\tilde{\boldsymbol{\kappa}} - (n+1)\Delta\kappa\right] \qquad j = 1,2,\ldots,\min(r,s); \quad n = 0,1,\ldots,N-1 \tag{2.30}$$

which is the same as:

$$S_{rs}^{g}(n\Delta\kappa) \begin{cases} 0 & n = 0 \\ \sum_{j=1}^{\min(r,s)} H_{rj}^{*}\left((n-1)\Delta\kappa + \frac{j}{m}\Delta\kappa\right) \cdot H_{sj}\left((n-1)\Delta\kappa + \frac{j}{m}\Delta\kappa\right) & n = 1,2,\ldots,N \end{cases} \tag{2.31}$$

2.4 Ergodicity of the generated samples

Equation (2.29) describes all terms of the CSDM (for all combinations of c and d) of a generated sample. This is the proof that each and all samples have the same CSDM and, therefore, the same auto- and cross-correlation functions. This proves the ergotic property of random Gaussian fields with finite period when are generated as described in Sec. 2.2.2.

For the special case when $\Delta\kappa$ goes to zero, Deodatis (1996) had already proved that the generated samples are ergodic. Herein, it will be shown that the two proofs are consistent, and that Eq. (2.31) yields the same results obtained by Deodatis when $\Delta\kappa$ goes to zero. In particular, it is shown that in such special case, the *actual* spectrum of each generated

sample tends to the *target* spectrum:

$$\begin{aligned}
\lim_{\Delta\kappa\to 0} S^g_{rs}(l\Delta\kappa) &= \begin{cases} 0 & l=0 \\ \lim\limits_{\Delta\kappa\to 0} \sum\limits_{j=1}^{\min(r,s)} H^*_{rj}\left((l-1)\Delta\kappa+\dfrac{j}{m}\Delta\kappa\right) & \\ \qquad \cdot H_{sj}\left((l-1)\Delta\kappa+\dfrac{j}{m}\Delta\kappa\right); & l=1,2,\ldots,N \end{cases} \\
&= \begin{cases} 0 & \text{when} \quad l=0 \\ \sum_{j=1}^{\min(r,s)} H^*_{rj}(\kappa)\cdot H_{sj}(\kappa); & \text{when} \quad l=1,2,\ldots,N \end{cases} \\
&= \begin{cases} 0 & \text{when} \quad l=0 \\ S_{rs}(\kappa) & \text{when} \quad l=1,2,\ldots,N \end{cases}
\end{aligned} \tag{2.32}$$

The proof of Eq. (2.32) is in the Appendix A. From Eq. (2.32) and for $r,s=1,\ldots,m$ we obtain $\mathbf{S}^g(\kappa)=\mathbf{S}(\kappa)$, which is what Deodatis proved considering the correlations (Deodatis, 1996). Therefore, Eq. (2.29) can be considered a generalization of Deodatis' proof of ergodicity, which is now extended also to the case of finite $\Delta\kappa$.

2.5 Numerical Examples

Three examples with different CSDM are presented in order to showcase the analytical results. To predict the CSDM of the samples, the analyst must first know the target CSDM and its resolution. For all examples, the closed-form expression, which allows one to compute the *actual* CSDM a priori, is validated. Additionally, rather than using ensemble averaging, the resulting CSDM is computed from a single sample, leveraging the ergodicity of the generated samples, also for finite $\Delta\kappa$.

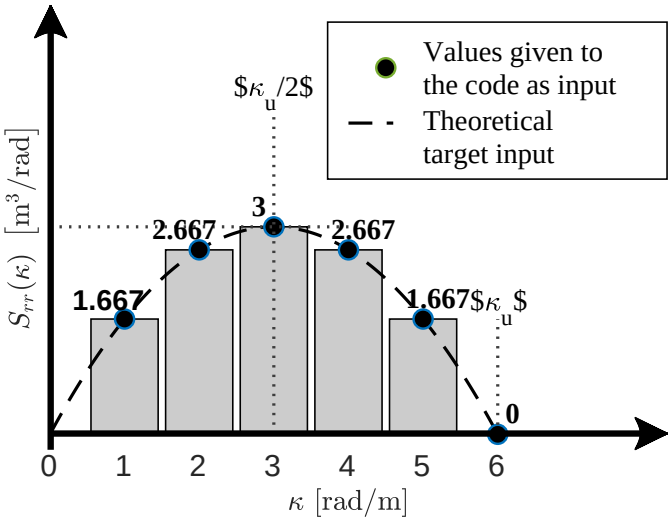


Figure 2.1: The values given to the code as target are $S\left(\frac{\Delta\kappa}{m}\right) = [1,2,3,2,1,0]'$.

2.5.1 Parabolic Spectrum

The first example consists of a one-dimensional three-variate vector field representing a displacement, with the following parabolic *target* spectrum: $N = 2$; $\kappa_u = 6$ rad/m; $\Delta\kappa = \frac{\kappa_u}{N} = 3$ rad/m; $M = 2N = 4$; $m = 3$ and

$$S_{rr}(\kappa) = 2\kappa - 2\frac{\kappa^2}{\kappa_u} \qquad \forall r \in [1,m] \tag{2.33}$$

$$\gamma_{rs}(\kappa) = 0.4 \qquad \forall r \in [2,m];\ s \in [1,r-1] \tag{2.34}$$

The number of discretization points N is too low to get Gaussianity, but for the scope of this paper, this is irrelevant, and ergodicity is guaranteed nonetheless.

The continuous triangular spectrum defined over κ cannot be directly used as input, rather its values computed at frequencies $\bar{\kappa}$ are required by the simulation methodology. The spectrum represented in Fig. 2.1 is the one actually used as input in the code.

The produced samples have $mM = 12$ discrete points. The empirical SDF computed using the generated samples has 12 points as well. However, only the first half of the points ($\kappa = 1, 2, \ldots, 6$) is meaningful. The second half is artificially introduced by the DFT because of the assumption of periodicity and is a mirror image of the first half.

2.5.2 Triangular Spectrum

The first example consists of a one-dimensional three-variate vector field representing a displacement, with the following characteristics: $N = 2$; $\kappa_u = 6$ rad/m; $\Delta\kappa = \frac{\kappa_u}{N} = 3$ rad/m; $M = 2N = 4$; $m = 3$, and

$$S_{rr}(\kappa) = \begin{cases} \kappa, & 0 \leq \kappa \leq \frac{\kappa_u}{2} \\ \kappa_u - \kappa, & \frac{\kappa_u}{2} < \kappa \leq \kappa_u \end{cases} \qquad \forall r \in [1, m] \tag{2.35}$$

$$\gamma_{rs}(\kappa) = 0.4 \qquad \forall r \in [2, m];\ s \in [1, r-1] \tag{2.36}$$

The number of discretization points N is too low to get Gaussianity, but for the scope of this paper, this is irrelevant, and ergodicity is guaranteed nonetheless.

The continuous triangular spectrum defined over κ cannot be directly used as input, rather its values computed at frequencies $\bar{\boldsymbol{\kappa}}$ are required by the simulation methodology. The spectrum represented in Fig. 2.1 is the one actually used as input in the code.

The produced samples have $mM = 12$ discrete points. The empirical SDF computed using the generated samples has 12 points as well. However, only the first half of the points ($\kappa = 1, 2, \ldots, 6$) is meaningful. The second half is artificially introduced by the DFT because of the assumption of periodicity and is a mirror image of the first half. For instance, S_{22} and S_{33} are reported in Fig. 2.2.

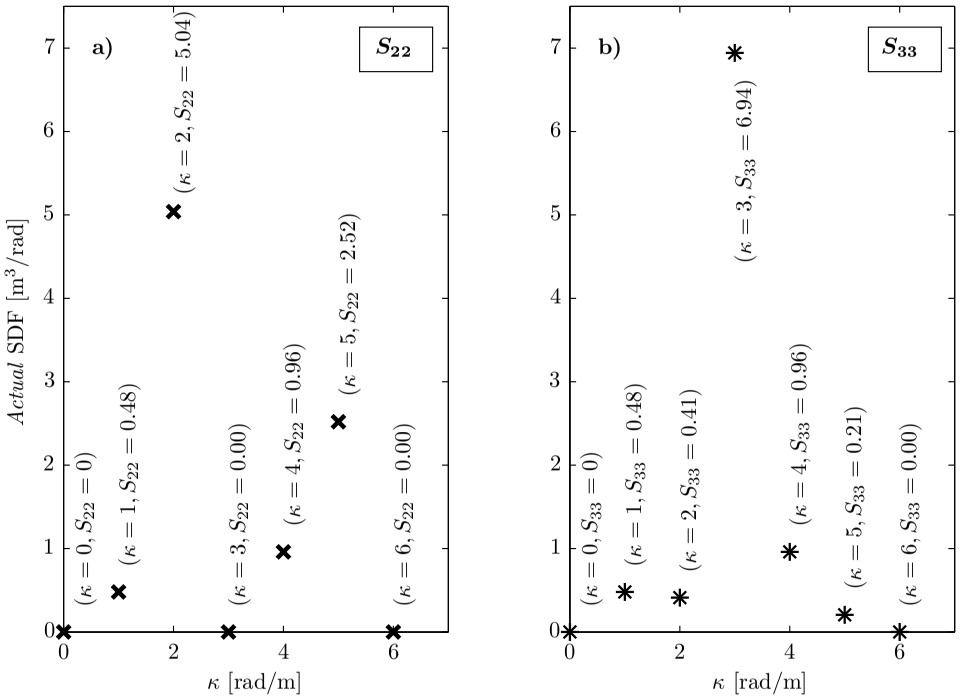


Figure 2.2: Triangular spectra of the second (a) and third (b) components of the vector process computed by the empirical autospectrum presented by Eq. (2.14).

κ	*Target*	*Actual* S_{22}	*Actual* S_{33}
$1 \cup 2 \cup 3$	1+2+3=**6**	0.48+5.04+0=**5.52**	0.48+0.41+6.94=**7.83**
$4 \cup 5 \cup 6$	2+1+0=**3**	0.96+2.52+0=**3.48**	0.96+0.21+0=**1.17**

Table 2.1: Triangular spectrum in m^3/rad. The *actual* spectra do not match the *target*, neither in an average sense.

It is evident that the empirical spectra do not match the target. This could be attributed to the fact that the "resolution" initially introduced is not $\frac{\Delta\kappa}{m}$ but $\Delta\kappa$. However, even when averaging (if we consider the values) or summing (if we consider the areas) the results in intervals of width $\Delta\kappa$, the resulting spectra still do not match the target (see Table 2.1 and Figs. 2.3–2.4).

In Sec. 2.4, we saw that when $\Delta\kappa$ tends to zero, the CSDM of the samples tends to the target. In fact, increasing the number of intervals N, that is reducing $\Delta\kappa$, the CSDM computed with resolution $\Delta\kappa$ becomes closer to the target, as Fig. 2.5 shows. However, no matter how small $\Delta\kappa$ is, if we compute the CSDM with a resolution of $\frac{\Delta\kappa}{m}$, we see that it is scattered (see Fig. 2.6). Figures 2.5 and 2.6 clearly show that the most meaningful way to compute the CSDM is to consider intervals of width $\Delta\kappa$, as will be done in the remainder of the paper.

The difference between the *actual* spectra of the produced samples and the *target* remains finite even when increasing N, as the detail in Fig. 2.5 shows. Figures 2.7–2.9 represent the difference between the *actual* spectrum and the *target* triangular SDF for different discretizations. As already said, the difference is due to the fact that each point of the *actual* spectra of the samples is a combination of the values of the *target* spectra at m points over the interval $\Delta\kappa$. Therefore, this difference increases with the difference between the values of the *target* spectra over one interval of width $\Delta\kappa$. We can conclude that the difference is larger when $\Delta\kappa$ is larger and when the derivative of the *target* autospectrum is larger. In fact, for a linear (triangular) spectrum, the difference takes constant values, while the next

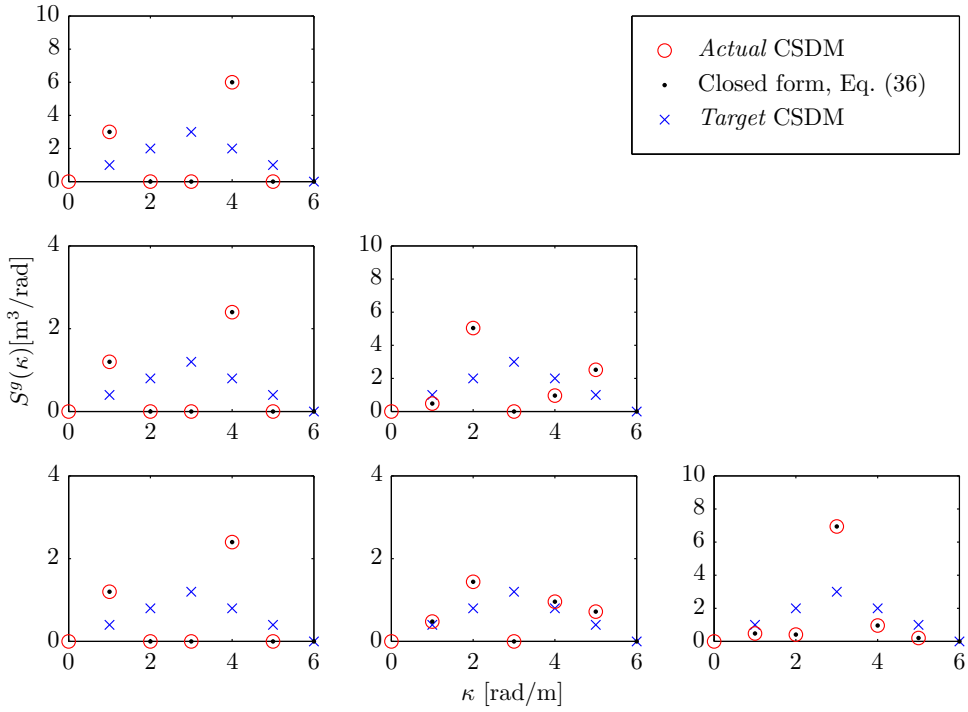


Figure 2.3: Triangular spectrum, cross-spectral density matrix. The "*Actual* CSDM" is the empirical spectrum computed applying Eq. (2.14) to one produced sample; the "Closed form" represents the results of Eq. (2.29); the "*Target* CSDM" is the input given to the simulation algorithm.

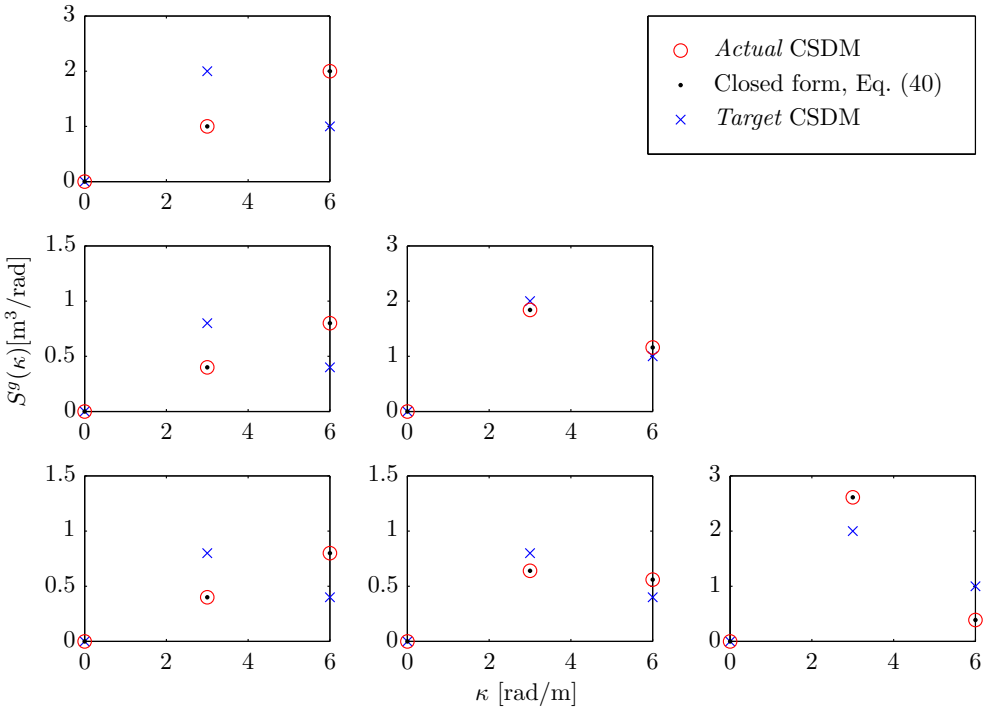


Figure 2.4: Triangular spectrum, cross-spectral density matrix. Data are computed as in Fig. 2.3, but at intervals of width $\Delta\kappa$, rather then $\frac{\Delta\kappa}{m}$. Each value in this plot is the sum of m values in Fig. 2.3. It is clear that also with this resolution, the closed-form expression perfectly describes the spectrum of the produced samples but it is different from the target.

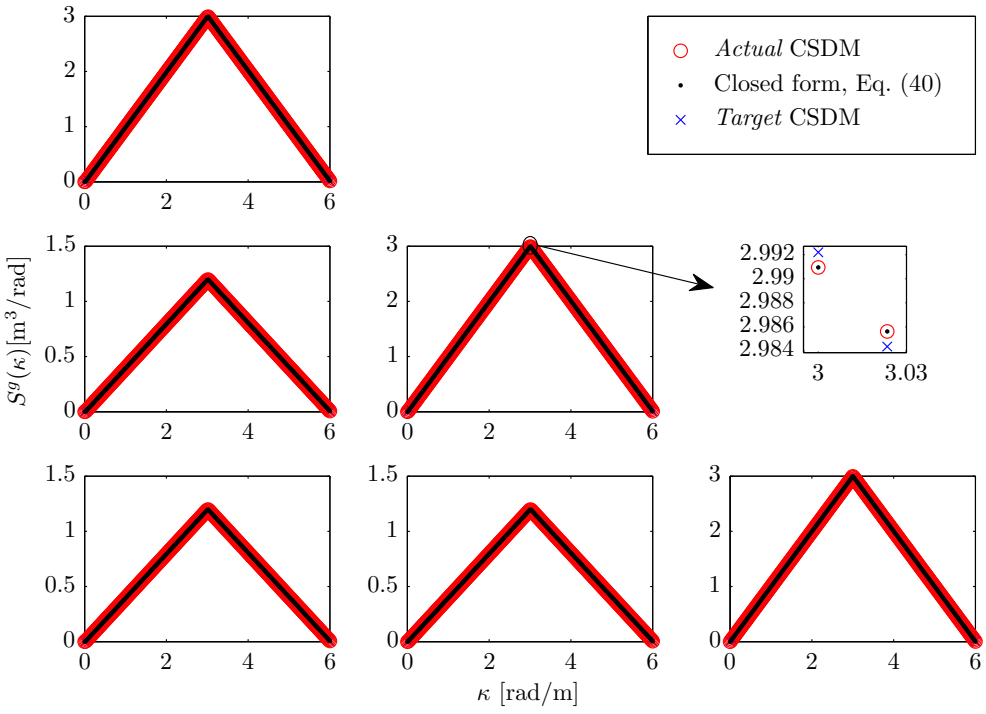


Figure 2.5: Triangular spectrum, cross-spectral density matrix. Data are computed as in Fig. 2.4. N has been set equal to 256, therefore $\Delta\kappa = \frac{6}{256} \cong 0.0234$ rad/m. The *actual* CSDM is much closer to the *target* than the one in Fig. 2.4.

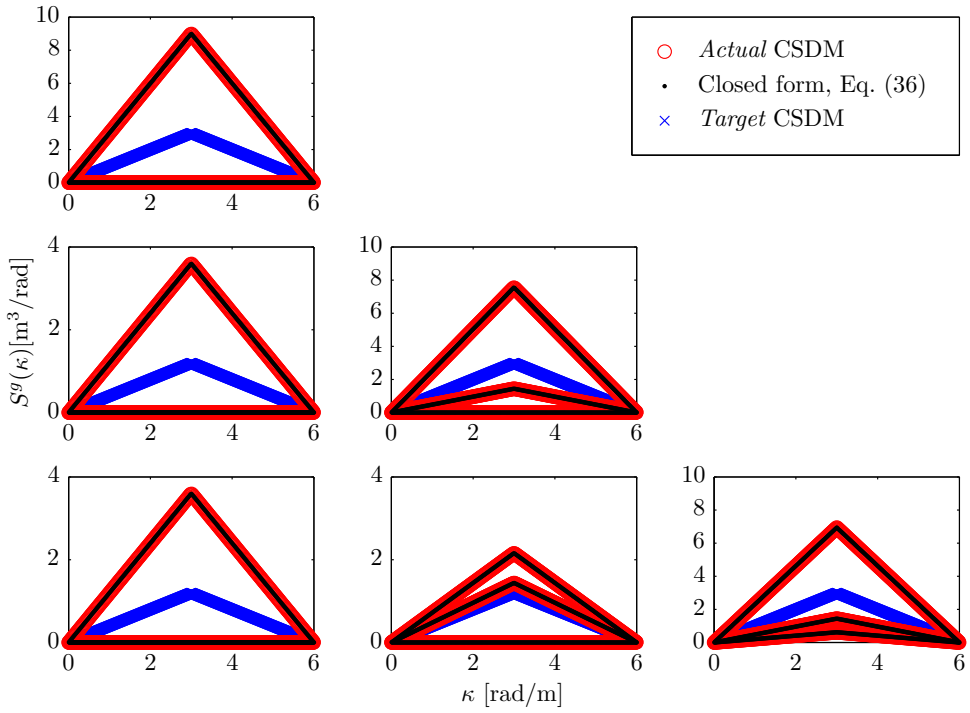


Figure 2.6: Triangular spectrum, cross-spectral density matrix. Data are computed as in Fig. 2.3. N has been set equal to 256, therefore $\frac{\Delta\kappa}{m} = \frac{6}{3\cdot 256} \cong 0.0078$ rad/m. Measuring the CSDM at intervals of $\frac{\Delta\kappa}{m}$, the *actual* CSDM is always scattered, as long as $\Delta\kappa$ is finite.

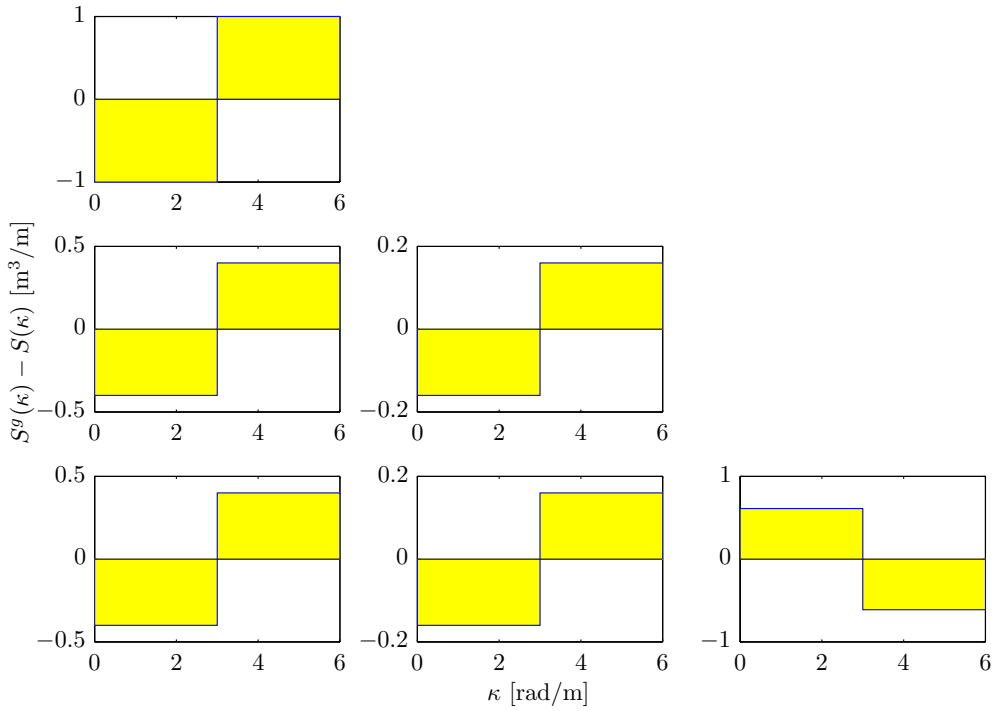


Figure 2.7: Triangular spectrum, difference between the *actual* CSDM computed as in Fig. 2.4 and the *target*. N has been set equal to 2.

numerical example will show that for a parabolic *target* spectrum, the difference is a linear function.

It is evident that the difference between the values of the *target* spectra and the *actual* spectra are always finite, and can be computed a priori by means of Eqs. (2.29) for a specified CSDM and selected interval $\Delta\kappa$.

2.5.3 Parabolic Spectrum

The second example consists of a one-dimensional three-variate vector field representing a displacement, with the following parabolic *target* spectrum: $N = 2$; $\kappa_u = 6$ rad/m;

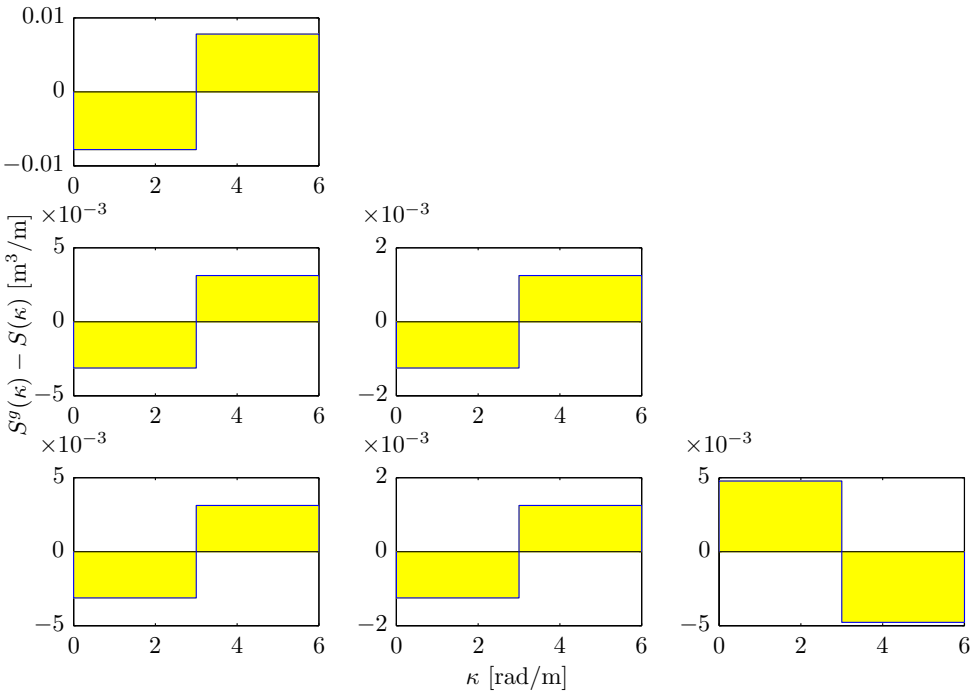


Figure 2.8: Triangular spectrum, difference between the *actual* CSDM computed as in Fig. 2.4 and the *target*. N has been set equal to 256. The absolute value of the difference is constant, because the derivatives of the SDF's are constant.

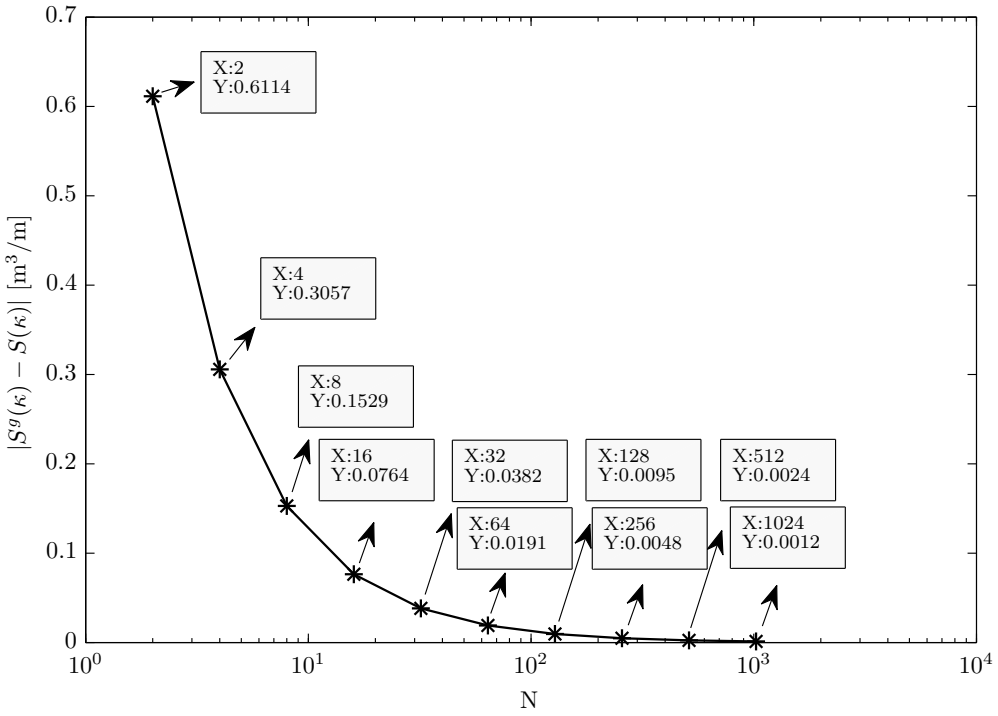


Figure 2.9: Triangular spectrum, difference between the *actual* CSDM computed as in Fig. 2.4 and the *target* for different values of N. This difference is plotted only for component S_{33}. The absolute value of the difference is constant over κ, as shown by Figs. 2.7–2.8 and it decreases as N increases (i.e., as $\Delta\kappa$ decreases). Similar differences could be obtained for the other components of the CSDM.

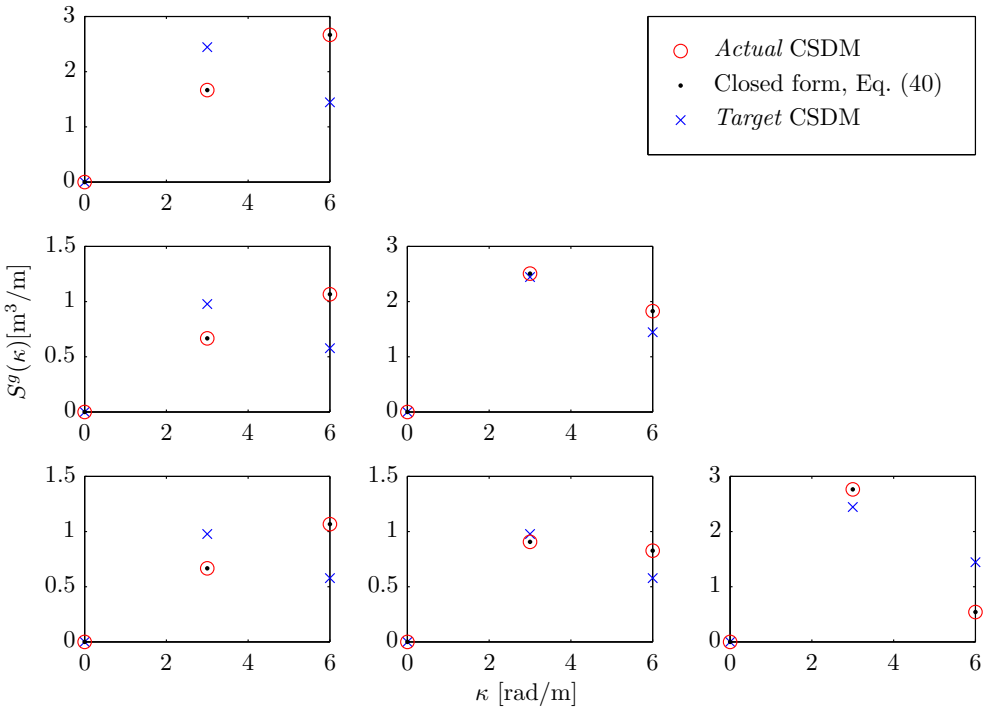


Figure 2.10: Parabolic spectrum, cross-spectral density matrix. The "*Actual* CSDM" has been computed applying Eq. (2.14) to one produced sample; the "Closed form" represents the results of Eq. (2.29); the "*Target* CSDM" is the input given to the simulation algorithm. The values are computed at intervals of width $\Delta\kappa$ and N is equal to 2.

$\Delta\kappa = \frac{\kappa_u}{N} = 3$ rad/m; $\quad M = 2N = 4$; $\quad m = 3$ and

$$S_{rr}(\kappa) = 2\kappa - 2\frac{\kappa^2}{\kappa_u} \qquad \forall r \in [1,m] \tag{2.37}$$

$$\gamma_{rs}(\kappa) = 0.4 \qquad \forall r \in [2,m];\ s \in [1,r-1] \tag{2.38}$$

Figures 2.10 and 2.11 represent the *actual* CSDM of the produced samples for $N = 2$ and $N = 64$, respectively. In this case the difference between the *target* and the *actual* spectrum is linear because the derivatives of the SDFs are linear (Fig. 2.12) .

For this example, we have computed also the variances of the produced samples. Ac-

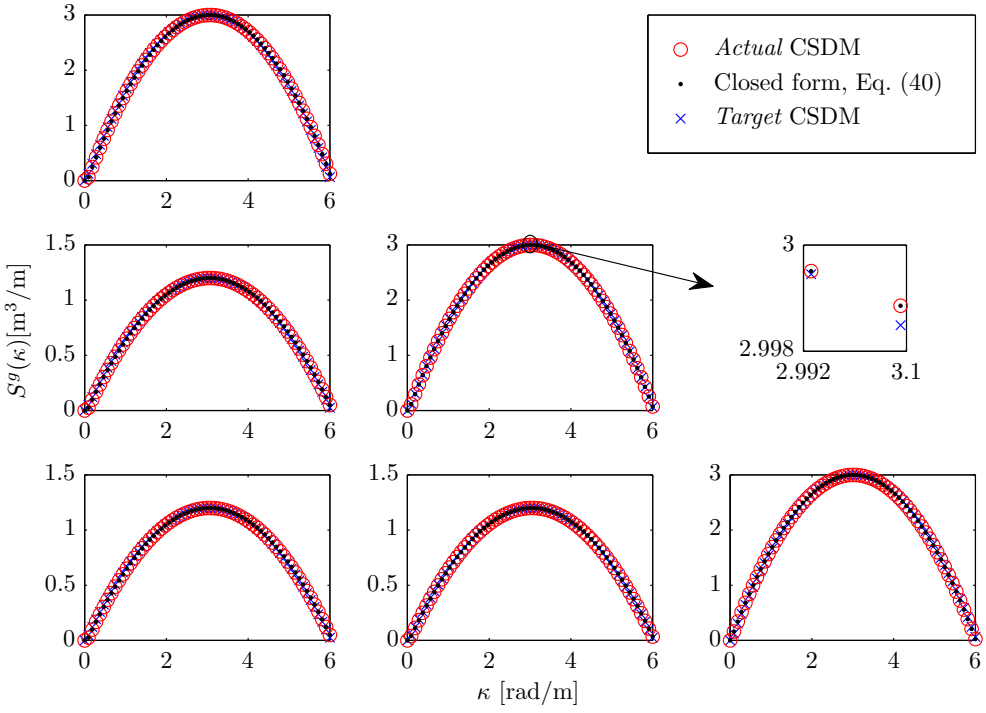


Figure 2.11: Parabolic spectrum, cross-spectral density matrix. The "*Actual* CSDM" has been computed applying Eq. (2.14) to one produced sample; the "Closed form" expression represents the results of Eq. (2.29); the "*Target* CSDM" is the input given to the simulation algorithm. The values are computed at intervals of width $\Delta\kappa$ and N is equal to 64.

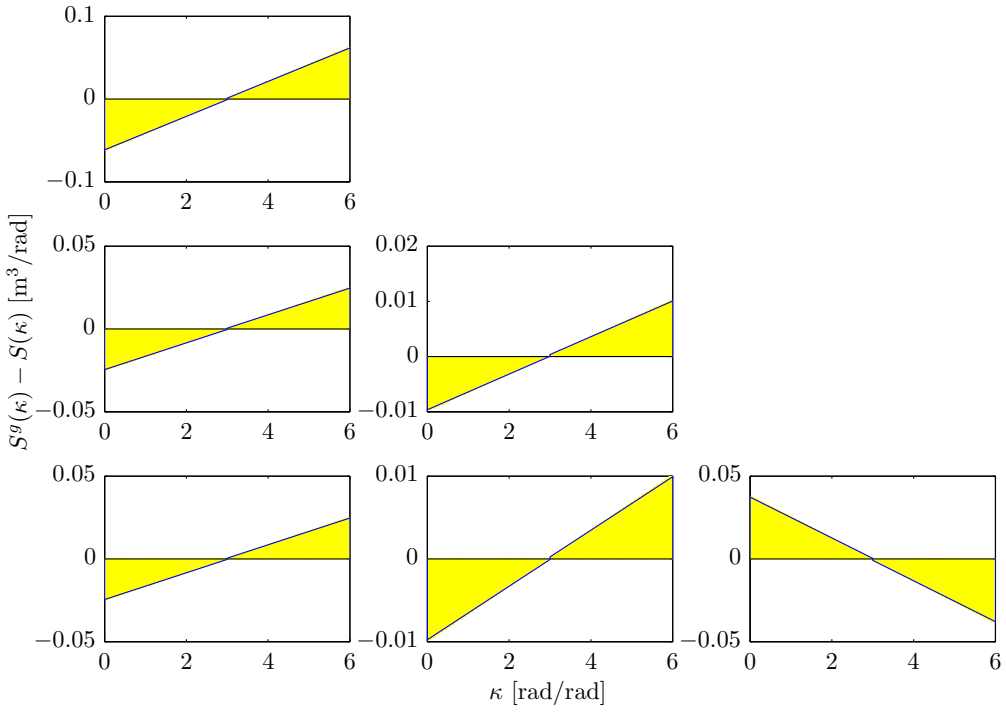


Figure 2.12: Parabolic spectrum, difference between the *actual* CSDM computed as in Fig. 2.4 and the *target*. N has been set equal to 64. The difference is a linear function, because the derivatives of the SDF's are linear.

cording to the parabolic auto-spectra, the variance of all components should be

$$\sigma_r^2 = 2\int_0^{\kappa_u}\left(2\kappa - 2\frac{\kappa^2}{\kappa_u}\right)\mathrm{d}\kappa = 24\,\mathrm{m}^2; \qquad r = 1,2,3 \tag{2.39}$$

If we numerically compute the variances of the produced samples we get

$$\sigma_1^2 = 28.4\,\mathrm{m}^2; \quad \sigma_2^2 = 28.4\,\mathrm{m}^2; \quad \sigma_3^2 = 21.6\,\mathrm{m}^2 \tag{2.40}$$

These results are different from the theoretical variances in part because $N = 2$ yields a very coarse discretization of the wave number domain, so the *target* spectra given as input to the code are not capturing well the theoretical spectra. However, the three variances also differ from each other, even if the three *target* auto-spectra are the same, and this is due to the fact that the *actual* auto-spectra depend also on the *target* coherences, as Eq. (2.30) indicates. To the best of our knowledge, this aspect has never been pointed out.

2.5.4 Kaimal Spectrum

Finally, the Kaimal spectrum with Davenport coherence used by Deodatis (1996) is used for the simulation of a one-dimensional three-variate random process. The number of intervals is set equal to $N = 2048$ and the cut-off frequency equal to $\kappa_u = 6$ rad/s. In this case, the difference between the *target* and the *actual* spectrum is very small but not zero (Fig. 2.13). Also in this case, the closed form Eq. (2.29), accurately describes the resulting *actual* CSDM.

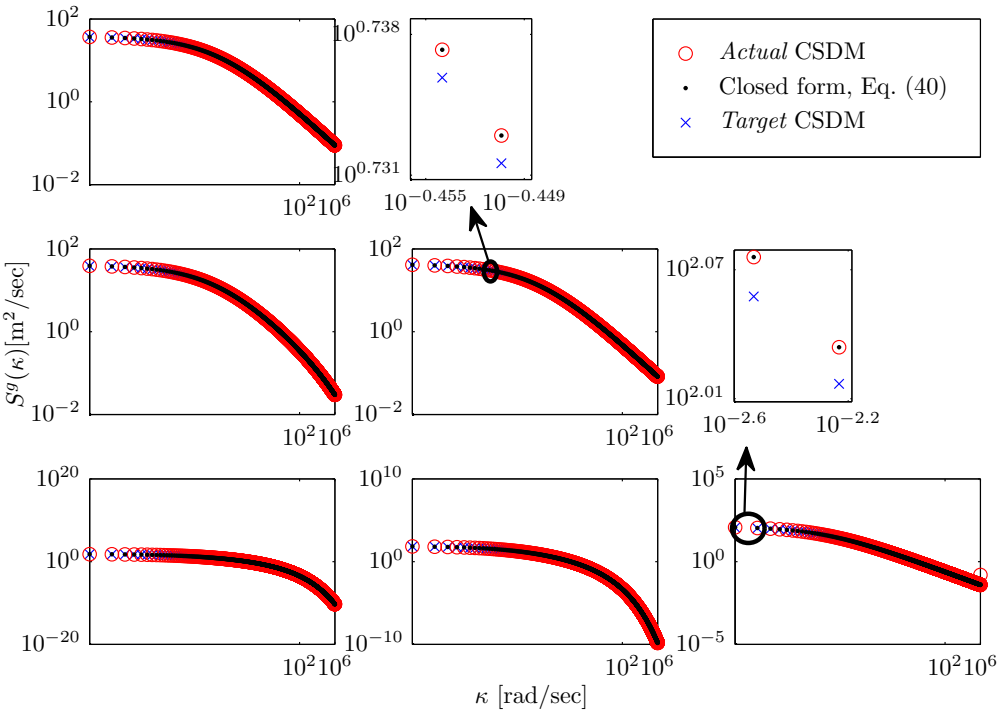


Figure 2.13: Kaimal spectrum, cross-spectral density matrix. The "*Actual* CSDM" has been computed applying Eq. (2.14) to one produced sample; the "Closed form" expression represents the results of Eq. (2.29); the "*Target* CSDM" is the input given to the simulation algorithm. The values are computed at intervals of width $\Delta\kappa$ and N is equal to 2048.

2.5.5 Triangular spectrum with complex CSDM

This example consists of a one-dimensional three-variate vector field with the following characteristics: $N = 16$; $\kappa_u = 6$ rad/m; $\Delta\kappa = \frac{\kappa_u}{N} = 0.3750$ rad/m; $M = 2N = 32$; $m = 4$ and

$$S_{rr}(\kappa) = \begin{cases} \kappa, & 0 \leq \kappa \leq \frac{\kappa_u}{2} \\ \kappa_u - \kappa, & \frac{\kappa_u}{2} < \kappa \leq \kappa_u \end{cases} \qquad \forall r \in [1, m] \tag{2.41}$$

$$\gamma_{rs}(\kappa) = 0.4 + 0.4\mathrm{i} \qquad \forall r \in [2, m]; s \in [1, r-1] \tag{2.42}$$

Figures 2.14 and 2.15 show the real and imaginary parts of the CSDM of the produced samples at wave numbers with intervals $\frac{\Delta\kappa}{m}$. The different contributions at the various secondary frequencies in the real part of the *actual* spectrum can be noticed in Fig. 2.14. The contributions to the spectrum at some secondary frequencies of the off-diagonal terms may take negative values, something which would not be observed if the same example excluded the imaginary part of the CSDM. Additionally, the imaginary part of the CSDM in Fig. 2.15 takes values different from 0 only at frequencies $\tilde{\boldsymbol{\kappa}}$.

Figures 2.16 and 2.17 depict the real and imaginary parts of the CSDM computed at intervals $\Delta\kappa$. Both figures validate that the closed form expression in Eq. (2.29) and the *actual* CSDM are in agreement when a complex CSDM is considered.

2.6 Concluding remarks

When the Spectral Representation Method is used together with the Frequency Double Indexing technique to produce multi-variate ergodic Gaussian samples, the resulting CSDM does not match exactly the *input* spectrum. The difference is generally small, and can be

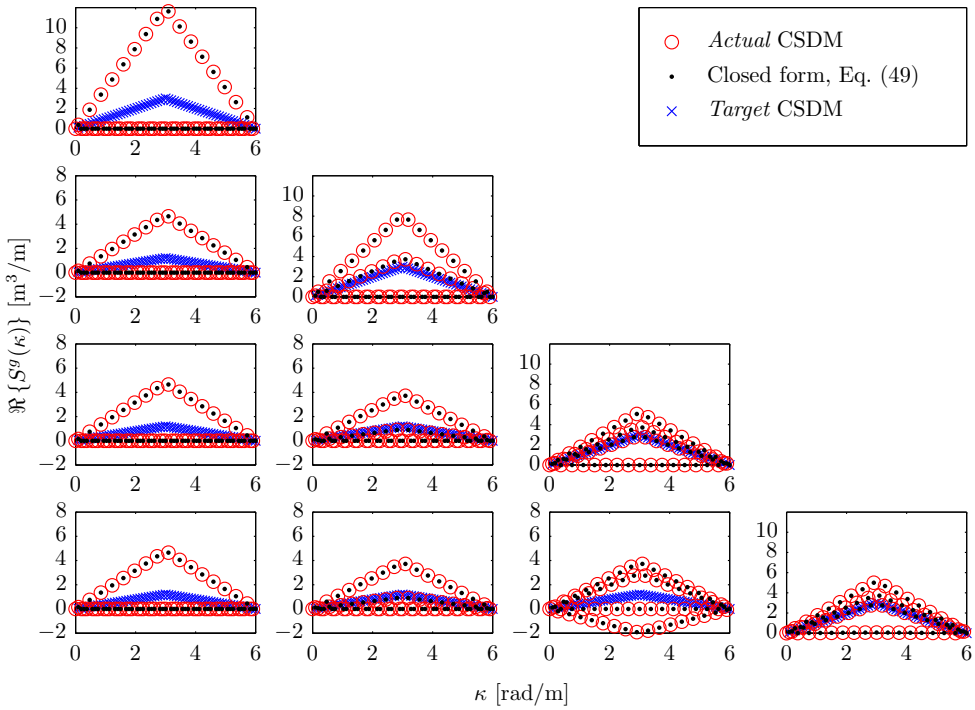


Figure 2.14: Real part of the complex cross-spectral density matrix. The "*Actual* CSDM" is the empirical spectrum computed applying Eq. (2.14) to one produced sample; the "Closed form" represents the results of Eq. (2.29); the "*Target* CDSM" is the input given to the simulation algorithm. The values are computed at intervals of width $\frac{\Delta\kappa}{m} = \frac{6}{4 \cdot 16}$ and N is equal to 16.

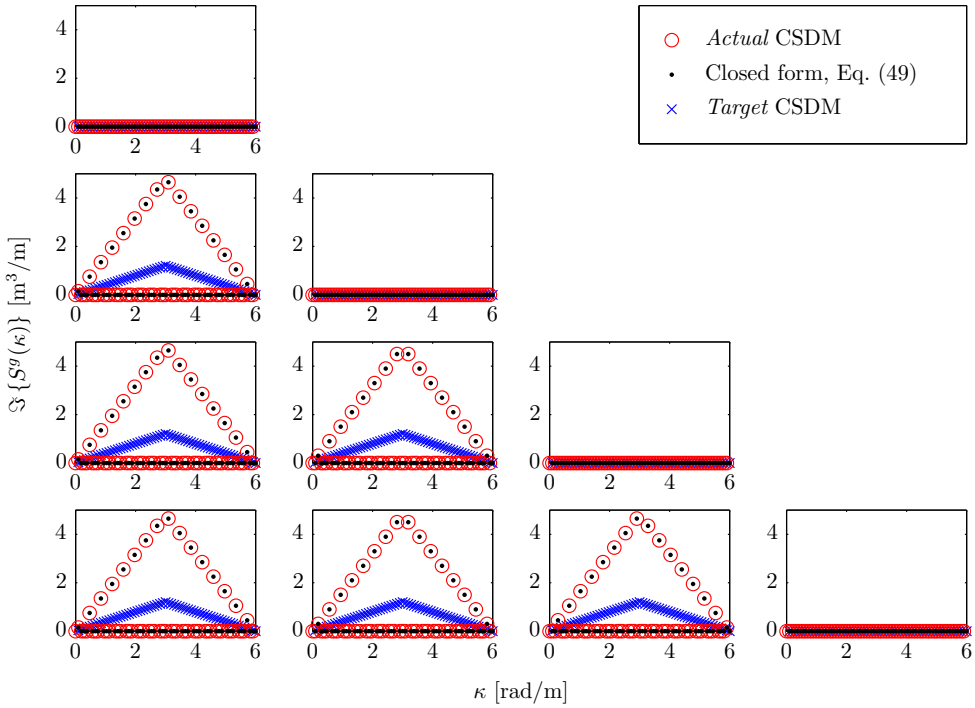


Figure 2.15: Imaginary part of the complex cross-spectral density matrix. The "*Actual* CSDM" is computed applying Eq. (2.14) to one produced sample; the "Closed-form" expression represents the results of Eq. (2.29); the "*Target* CDSM" is the input given to the simulation algorithm. The values are computed at intervals of width $\frac{\Delta\kappa}{m} = \frac{6}{4 \cdot 16}$ and N is equal to 16.

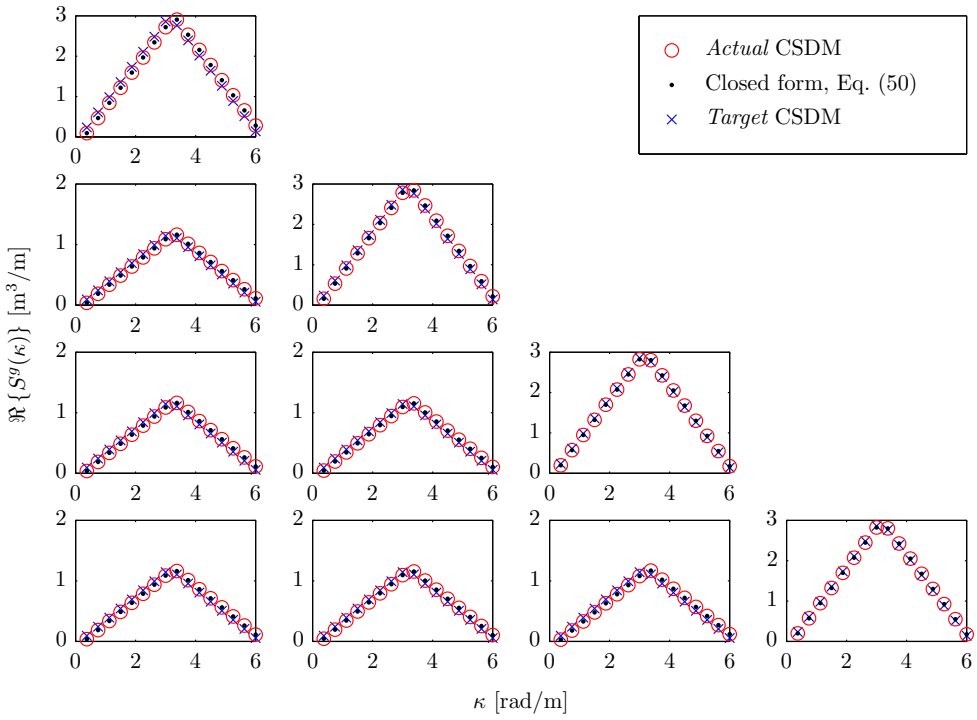


Figure 2.16: Real part of the complex cross-spectral density matrix. The "*Actual* CSDM" is computed applying Eq. (2.14) to one produced sample; the "Closed form" represents the results of Eq. (2.29); the "*Target* CDSM" is the input given to the simulation algorithm. The values are computed at intervals of width $\Delta\kappa$ and N is equal to 16.

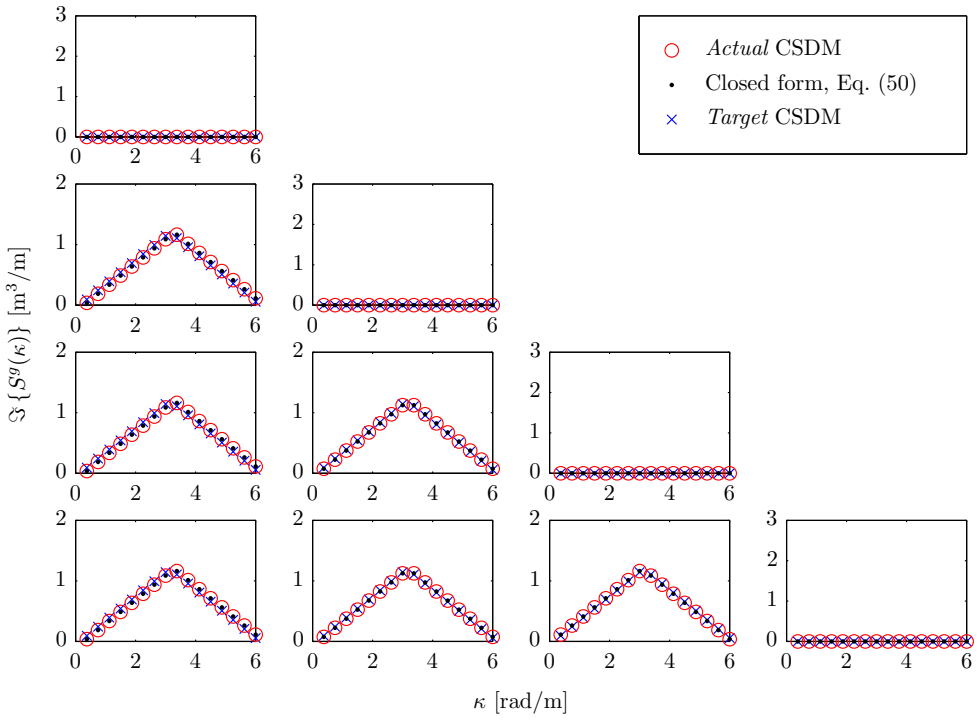


Figure 2.17: Imaginary part of the complex cross-spectral density matrix. The "*Actual* CSDM" has been computed applying Eq. (2.14) to one produced sample; the "Closed form" represents the results of Eq. (2.29); the "*Target* CDSM" is the input given to the simulation algorithm. The values are computed at intervals of width $\Delta\kappa$ and N is equal to 16.

easily reduced increasing the number of intervals N (i.e., reducing their width $\Delta\kappa$).

A closed-form expression that allows to compute a priori the *actual* CSDM of the produced samples has been provided. This is considered a generalization of the proof of ergodicity, which is shown to be valid also for finite $\Delta\kappa$.

The marginal probabilities of the various components of the produced samples are Gaussian for large N, but the variances are not equal to the integrals of the autocorrelations. These findings are particularly important when the technique is utilized to generate samples whose spectrum concentrates most of their energy at very low frequencies as for example in the case of the Kaimal spectrum used to capture the wind velocity fluctuations.

Chapter 3

Optimal representation of multi-dimensional random fields with a moderate number of samples: application to stochastic mechanics

3.1 Introductory remarks

In stochastic engineering problems, the proper consideration of the input variability is crucial to obtain an accurate and reliable solution. A large number of these problems involves uncertain quantities which should be modeled as multi-dimensional random fields. The use of multi-dimensional random fields gained momentum due to the continued increase in available computantional resourses and nowadays is commonly used in many disciplines. Several examples can be found in various fields of engineering. For instance, in structural engineering Christou and Bocchini (2014a) modeled the spatial distribution of corrosion over the upper flange of a steel I-beam as a two dimensional random field. Papadopoulos and Papadrakakis (2005) used two-dimensional uni-variate (2D-1V) stochastic fields to describe the non-homogeneous characteristics of initial imperfections in manufactured shells. In geotechnical engineering Popescu et al. Popescu et al. (2005) used two-dimensional fields to model the spatial variability of the soil mechanical characteristics. Similarly, in naval engineering Teixeira and Soares (2008) used two-dimensional fields to model the spatial corrosion propagation in ship-hull plates and computed their collapse strength.

The solution of these engineering problems is often obtained though simulation-based techniques, which are the most commonly used among the procedures available in the lit-

erature. Monte Carlo Simulation (MCS) is still considered the most reliable and versatile numerical technique for the solution of engineering problems affected by uncertainty. The drawback of MCS remains the large computational cost that prevents its use for many applications. In particular, the number N of deterministic runs which can be actually performed is limited by the complexity of the problem at hand, the type of input and the available time and computational resources. In many cases, this number N is small, —e.g. in the range $[50 \sim 1000]$—, too small for the law of large numbers to apply. Such sample size may be sufficient for the assessment of low-order statistics (i.e., mean or standard deviation at most), but certainly not to capture more information on the probability distribution. Thus, the result of a plain MCS would not be acceptable when a model of the entire distribution is sought. In these cases, a probabilistic technique that can capture in the most effective way the space of sample realizations of the random function, given a pre-determined number of samples should be used.

Multiple techniques have addressed the issue of sampling random functions more effectively, compared to plain MCS. For engineering problems the currently most popular method was presented in Grigoriu (2009) and is called "Stochastic Reduced-Order Models" (SROM). Grigoriu used SROM to find statistics of the state of linear dynamic systems with random and deterministic properties subjected to Gaussian and non-Gaussian noise Grigoriu (2013, 2010). Mignolet and Soize (2008) utilized SROM for the determination of the response of geometrically nonlinear structural dynamic systems and Warner et al. (2013) employed it to approximate the natural frequencies and modes of uncertain dynamic systems. The basic idea of SROM methods is to consider an optimization problem where the objective function quantifies the discrepancy between the statistics of the SROM and the random function being modeled. It will be shown that the methodology presented in this paper is rooted in a similar idea, but the optimization problem will be formulated in a different way. A discussion of the differences between SROM and the proposed method-

ology is provided in Section 3.4.

As an alternative to SROM, Functional Quantization (FQ) is a novel technique, proven to provide optimal approximations of random functions using a pre-determined number N of representative samples (Luschgy and Pagès, 2002; Luschgy et al., 2004). Moreover, some authors have used FQ directly as a variance reduction technique (Pagès and Printems, 2005; Lejay and Reutenauer, 2012). FQ is characterized by two major differences, compared to MCS: (1) the representative samples from FQ are selected not entirely at random and (2) the representative samples from FQ are not equally weighted. A few techniques for the selection of optimal samples and computation of associated probabilities based in the FQ concept have been presented in the literature. Some of the best known quantization techniques were presented by Lushgy and Pagés under the "Quantizer Design" umbrella Luschgy et al. (2010). The Quantizer Design I yields optimal results, whereas Quantizer Design II, III and IV are sub-optimal, but they are characterized by improved computational efficiency. All these techniques rely on the use of Karhunen-Loéve expansions, and therefore, they have been demonstrated only on Gaussian random functions for which such expansion is readily available. Another class of techniques was proposed by Corlay and Pagès (2015). They have the appeal of connecting FQ with the very popular stratified sampling approach. The authors presented four different versions of the approach but as in the case of the Quantizer Design, these techniques have been applied only to one-dimensional Gaussian processes.

To overcome the limitations that affect the previously mentioned FQ techniques, in this paper, a recently developed methodology called "Functional Quantization by Infinite-Dimensional Centroidal Voronoi Tessellation" (FQ-IDCVT) is considered (Miranda and Bocchini, 2015b). The FQ-IDCVT technique has been successfully used for one-dimensional, non-Gaussian and non-stationary processes (Miranda and Bocchini, 2013) and it has been shown to work particularly well against the curse of dimensionality that arises in stochastic

problems that use random functions for input quantities.

To further enhance the versatility of the technique, this paper presents it's extension to the case of random functions defined over a multi-dimensional domain. A description of the modifications to the existing algorithm required by the extension to multi-dimensional fields is accompanied by its demonstration on two numerical applications. The first example involves a two-dimensional Lognormal field, generated through the Spectral Representation method (SRM) (Shinozuka and Deodatis, 1996). For the simulation of the two dimensional field, a recently developed algorithm that approximates a non-Gaussian stationary multi-dimensional field is utilized (Christou and Bocchini, 2014a). Next, a sensitivity analysis is used to discuss the limits of applicability of the proposed approach and comments on the computational challenges associated with the extension to two-dimensional fields are provided. Finally, a second numerical application involving a two-dimensional panel in plane-stress with uncertain Young modulus, modeled as a two-dimensional stochastic field, is illustrated. For this case, the assessment of the effectiveness of the FQ-IDCVT technique is evaluated on the output quantities and compared to classical MCS.

3.2 Proposed approach

3.2.1 Functional Quantization

As already indicated, the goal of FQ is Grigoriu (2009, 2006); Stefanou and Papadrakakis (2007): to represent the probabilistic characteristics of a random function with a small-to-moderate number of carefully selected samples. Described in a sentence, the FQ approach consists in approximating a generic random function F by means of another random function F_N, which can be fully described (in a probabilistic sense) by a finite set of determin-

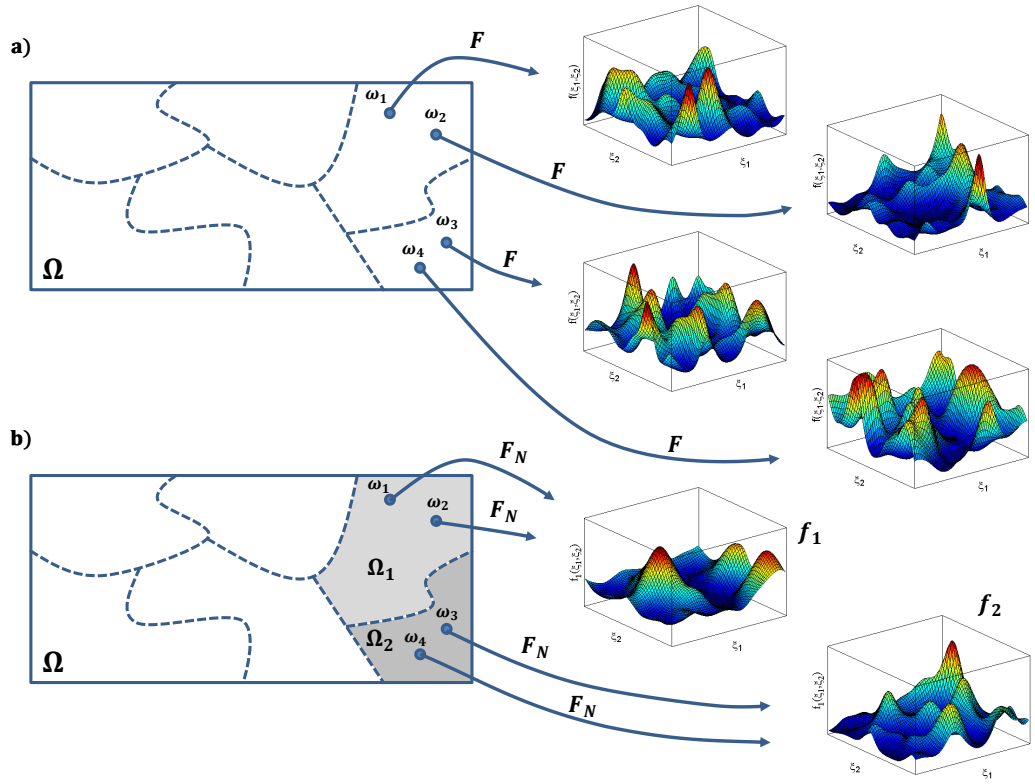


Figure 3.1: Upper part a) represents the random function F, the sample space Ω and the realizations that correspond to the single outcomes ω_i with $i = 1,2,3,4$. Bottom part b) shows how the random function F_N approximates F. The sample space Ω is partitioned into $\{\Omega_i\}_{i=1}^N$.Each subset Ω_i has a representative quantum and here two quanta that correspond to Ω_1 and Ω_2 are presented. Thus, every outcome that belongs to a certain subset Ω_i is associated with the same quantum. For example, the outcomes ω_1 and ω_2 that belong to Ω_1 are both associated with quantum f_1.

istic samples $\{f_i\}_{i=1}^N$ and associated probability masses $\{p_i\}_{i=1}^N$:

$$F \xrightarrow{\text{approximated by}} F_N \xrightarrow{\text{represented by}} \{f_i, p_i\}_{i=1}^N \tag{3.1}$$

For the sake of establishing a coherent nomenclature, these steps are briefly described in the following subsections.

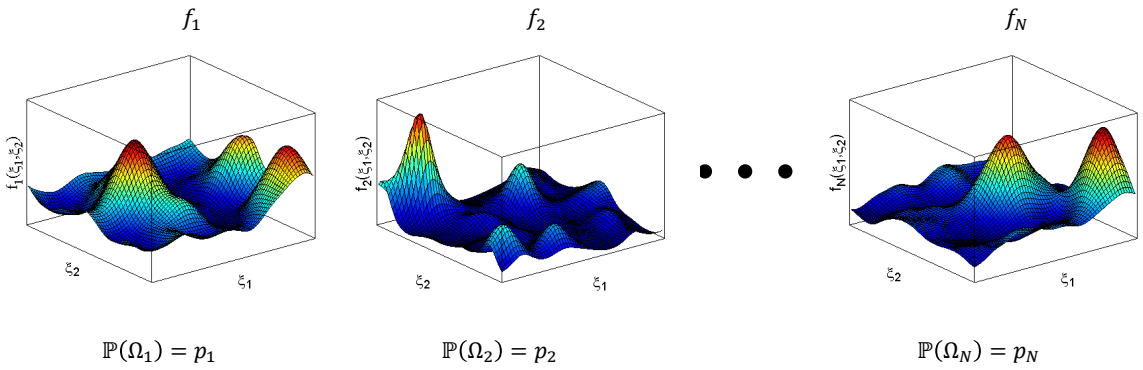


Figure 3.2: A quantizer for a two-dimensional random field.

3.2.1.1 Definition of the random function F

For a given probability space $(\Omega, \mathscr{F}, \mathbb{P})$, a bimeasurable random field $F(\boldsymbol{\xi}, \omega)$ is defined as:

$$F : \Xi \times \Omega \to \mathbb{R} \qquad \text{with } \Xi \subset \mathbb{R}^n \tag{3.2}$$

where Ξ is the spatial domain of interest in $\mathbb{R}^n$, Ω is the sample space and $\boldsymbol{\xi}$ is a point in Ξ. Therefore:

$$\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^n \tag{3.3}$$

A random function, as defined above, can also be interpreted as a random variable $F(\omega)$ with values in a certain function space which, for our purposes, can be assumed to be the space of square integrable functions $L^2(\Xi)$:

$$F : \Omega \to L^2(\Xi) \tag{3.4}$$

where $L^2(\Xi)$ is the space of square integrable functions.

The vast majority of random functions involved in realistic engineering problems are characterized by square-integrable realizations almost surely, finite mean and variance almost everywhere, and probability laws without any lumped masses. Therefore, the pre-

sented study focuses on this class of random fields.

3.2.1.2 Approximation of F by F_N

In the FQ approach, the random field F is approximated by another random field F_N taking a finite number N of "values" in $L^2(\Xi)$. F_N is defined by the following equation:

$$F_N(\boldsymbol{\xi}, \omega) = \sum_{i=1}^{N} f_i(\boldsymbol{\xi}) \cdot 1_{\Omega_i}(\omega) \tag{3.5}$$

where the deterministic functions $f_i(\boldsymbol{\xi})$ are called "quanta" and 1_{Ω_i} is the indicator function associated with the event Ω_i

$$1_{\Omega_i}(\omega) = \begin{cases} 1, & \text{if } \omega \in \Omega_i \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

The quanta $f_i(\boldsymbol{\xi})$ are defined in this paper as follows:

$$f_i(\boldsymbol{\xi}) = \begin{cases} \dfrac{\int_{\Omega_i} F(\omega)\, \mathrm{d}\mathbb{P}}{\mathbb{P}(\Omega_i)}, & \text{if } \mathbb{P}(\Omega_i) > 0, \\ \text{undefined}, & \text{otherwise} \end{cases} \tag{3.7}$$

where $\mathbb{P}(\Omega_i)$ is the probability associated with the subset Ω_i. Therefore, the sample space Ω is partitioned into $\{\Omega_i\}_{i=1}^{N}$ and each subset Ω_i has a representative quantum f_i, which is the average of all the sample functions associated with ω's which belong to the subset Ω_i. Then, F_N is defined by all the quanta and for a refined partition, it will closely approximate F. However, it should be noted that the infinite-dimensional function F (defined over an n-dimensional space) is approximated by a function F_N with a finite codomain of possible realizations. This may lead to the so called "curse of dimensionality", as it will be discussed in Section 3.5.1.1.

The partition of Ω can be also projected to the space of square integrable functions $L^2(\Xi)$, where the realizations of F and F_N lie. From this perspective, the $L^2(\Xi)$ space is tasseled into $\{V\}_{i=1}^N$, where each tassel V_i collects all the realizations $F(\omega)$ with $\omega \in \Omega_i$. This tessellation will be further explained in Section 3.2.3.

3.2.1.3 Representation of F_N by the quantizer

Figure 3.1a is a pictorial representation of the random function F mapping outcomes of the sample space Ω into realizations defined over the two-dimensional spatial domain Ξ. When the FQ technique is utilized, the sample space Ω is partitioned into subsets $\{\Omega_i\}_{i=1}^N$, and all outcomes belonging to the same subset Ω_i are mapped to the quantum f_i that represents that specific subset (Figure 3.1b).

According to Eq. (3.5) and Figure 3.1b, the random function F_N is fully described by the chosen partition of Ω and by the set of quanta f_i. In particular, for a probabilistic characterization of F_N, it is not even necessary to know the partition of Ω, but only the probability associated with each subset Ω_i. For this reason, the collection of deterministic functions $\{f_i\}_{i=1}^N$ and their associated probabilities $\{p_i\}_{i=1}^N$ (i.e., $\mathbb{P}(\Omega_i)$) is all that is necessary for the characterization of F_N and, in turn, for the approximate probabilistic description of F.

For the reminder of the paper, the set of quanta $\{f_i\}_{i=1}^N$ and associated probabilities $\{p_i\}_{i=1}^N$ is called the "quantizer", and the number N of quanta is called the "quantizer size". Figure 3.2 shows a representation of a two-dimensional quantizer.

3.2.1.4 Quantizers in engineering problems

When solving a problem in stochastic mechanics, the quanta can be used as inputs in a simulation-based fashion. In particular, having specified the $\{f_i, p_i\}_{i=1}^N$ pairs, the system response is obtained by running N deterministic analyses with each of the functions f_i used

as independent input. Then, the probabilistic characterization of the outcome is determined by weighing the result of each analysis with the associated probability p_i. For further theoretical details, the reader may refer to Luschgy and Pagès (2002) or Miranda and Bocchini (2015b).

The identification of an optimal partition of the sample space Ω is an optimization problem and different solutions could be obtained through the use of various objective functions. In this paper, optimality is sought by imposing the minimization of a mean square error function named "distortion". In general, any FQ technique has to perform the following operations:

- for a fixed N, optimally partition the sample space Ω into the events $\{\Omega_i\}_{i=1}^N$ and so find the corresponding tassels $\{V_i\}_{i=1}^N$ in the space of square integrable functions $L^2(\Xi)$
- obtain the representative deterministic functions $\{f_i\}_{i=1}^N$ for all subsets $\{\Omega_i\}_{i=1}^N$ and corresponding tassels $\{V_i\}_{i=1}^N$
- determine the probability masses $\mathbb{P}(\Omega_i) = \mathbb{P}_F(V_i) = p_i$ of the subsets $\{\Omega_i\}_{i=1}^N$ or the corresponding tassels $\{V_i\}_{i=1}^N$

The computational cost to obtain an optimal quantizer is not negligible. Nevertheless, it is usually small compared to the total time needed to perform a probabilistic analysis of a complex engineering problem.

3.2.2 Tessellation of Finite-Dimentional Spaces

An important characteristic of the FQ technique is its close conceptual relationship with the Voronoi Tessellation (VT) and the Centroidal Voronoi Tessellation (CVT) of an Euclidean

space $\mathbb{R}^n$. When the VT is applied, a finite-dimensional Euclidean space is partitioned in regions $\{V_i\}_{i=1}^N$, called "Voronoi tassels". Each one of these is an n–dimensional convex polyhedron with a generating point $\check{\mathbf{y}}_i \in \mathbb{R}^n$.

The tassels are defined in such a way that all the points $\mathbf{y} \in \mathbb{R}^n$ that belong to tassel V_i are closer to the generating point $\check{\mathbf{y}}_i$ than to any other point $\check{\mathbf{y}}_{j\neq i}$.

$$V_i = \Big\{ \mathbf{y} \in \mathbb{R}^n \quad | \quad \|\mathbf{y} - \check{\mathbf{y}}_i\| < \|\mathbf{y} - \check{\mathbf{y}}_j\| \\ \text{for } j = 1, 2, \ldots, N; j \neq i \Big\} \tag{3.8}$$

where $\|\cdot\|$ is the Euclidean norm.

The CVT of a finite-dimensional space is a special case of the VT where the generating point $\check{\mathbf{y}}_i$ of the tassel V_i is also the mass centroid of the convex region. The mass centroid can be defined as:

$$\bar{\mathbf{y}}_i = \frac{\int_{V_i} \mathbf{y} \cdot \rho(\mathbf{y}) \, \mathrm{d}V}{\int_{V_i} \rho(\mathbf{y}) \, \mathrm{d}V} \tag{3.9}$$

where $\rho(\mathbf{y})$ is the mass density at point $\mathbf{y}$. Therefore, to construct a CVT it is required that the generating points $\check{\mathbf{y}}_i$ used in Equation (3.8) coincide with the mass centroids $\bar{\mathbf{y}}_i$ as computed by Equation (3.9):

$$\check{\mathbf{y}}_i \equiv \bar{\mathbf{y}}_i \quad \forall i \tag{3.10}$$

3.2.3 Infinite-Dimensional Centroidal Voronoi Tessellation

Miranda and Bocchini Miranda and Bocchini (2015b) extended the concept of finite-dimensional VT and CVT to the infinite-dimensional Hilbert space of square-integrable functions $L^2(\Xi)$. As already mentioned, the random function $F(\boldsymbol{\xi}, \omega)$ can be seen as a random variable $F(\omega)$ with values in the Hilbert space where the tassels lie. Such tassels can be defined as follows:

$$V_i = \Big\{ F(\omega) \in L^2(\Xi) \quad | \quad \|F(\omega) - \check{f}_i\|_{L^2(\Xi)} < \|F(\omega) - \check{f}_j\|_{L^2(\Xi)} \quad \text{for } j = 1,2,\ldots,N; j \neq i \Big\} \tag{3.11}$$

where $\check{f}_i$ is the generating point of tassel V_i. In the Infinite-Dimensional Centroidal Voronoi Tessellation case (IDCVT), the generating points $\check{f}_i$ and the centroids $\bar{f}_i$ of tassel V_i coincide, where the centroids can be computed as follows:

$$\bar{f}_i = \frac{\int_{V_i} F(\omega)\, \mathrm{d}\mathbb{P}_F}{\mathbb{P}_F(V_i)} \tag{3.12}$$

in which $\mathbb{P}_F$ is defined by the probability law of the random function F and the integration is in the Bochner sense. In addition, it has been proved that such functions are also optimal quanta in the FQ sense Miranda and Bocchini (2015b):

$$\check{f}_i \equiv \bar{f}_i \equiv f_i \quad \forall i \tag{3.13}$$

When the tessellation of $L^2(\Xi)$ is defined, the probability masses $\mathbb{P}_F(V_i) = \mathbb{P}(\Omega_i) = p_i$ can be computed. Each tassel V_i is associated with a subset Ω_i of the sample space Ω. By this conceptual operation, almost all the space Ω is partitioned into $\{\Omega_i\}_{i=1}^N$. In other words, the partition of the sample space Ω is induced from the tessellation of $L^2(\Xi)$. Similarly, the probability masses $\mathbb{P}(\Omega_i)$ associated with the various $\{\Omega_i\}_{i=1}^N$ are determined applying the frequentist definition of probability to the tassels of $L^2(\Xi)$, as will be explained later in Equation (3.17).

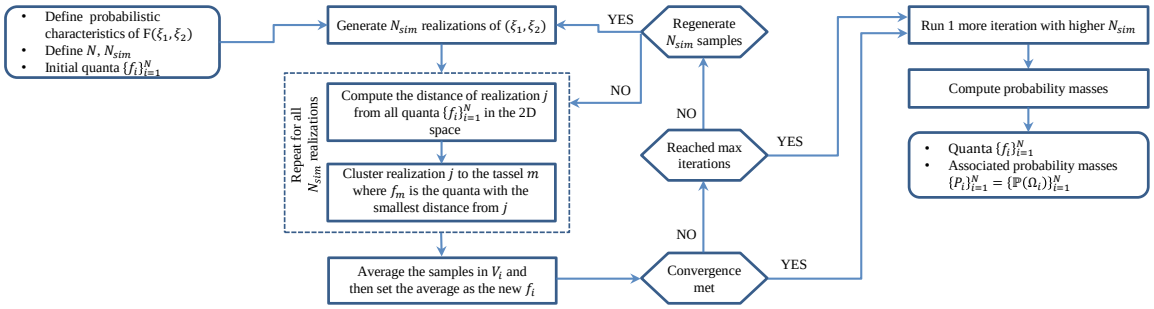


Figure 3.3: Main steps of the FQ algorithm.

3.3 Algorithm

The FQ-IDCVT technique is an extension of the iterative algorithm of Lloyd's Method Ju et al. (2002). Despite the complex underlying theoretical bases, the implementation of the algorithm is very simple. Figure 3.3 shows the flowchart of the FQ-IDCVT methodology for a two-dimensional uni-variate (2D-1V) random function and the basic steps can be summarized as follows.

1. Input data required for the FQ-IDCVT technique
   - quantizer size N
   - computational parameter $N_{sim} = N \cdot k$, with $k \in [100, 5000]$
   - data required to probabilistically define the random field $F(\xi_1, \xi_2)$, such as marginal distribution and spectral density
   - initial set of quanta $\{f_i\}_{i=1}^N$, for instance generated as N random samples
2. Simulate N_{sim} realizations of the random function F. The realizations are generated at the beginning of the algorithm and optionally they can be regenerated later. When the samples are regenerated within each iteration, the computational cost is increased and the convergence is delayed. However, the outcome of the technique is less sensitive to the seed selected for the first generation of N_{sim} samples.

3. Compute the distances of each realization of F from each quantum $\{f_i\}_{i=1}^N$ according to the $L^2(\Xi)$ norm.

4. Define a set of tassels V_i, $i = 1,2,\ldots N$, collecting in each V_i all the realizations of F that are closer to quantum f_i than to any other quantum f_j according to Equation (3.11).

5. Average the realizations in each tassel $\{V_i\}_{i=1}^N$ and set the average as the new quantum f_i of the respective tassel V_i using a discrete version of Equation (3.12):

$$f_i(\xi_1,\xi_2) = \frac{1}{N_i}\sum_{k=1}^{N_i} \hat{f}_k(\xi_1,\xi_2) \qquad \text{with } \hat{f}_k \in V_i \tag{3.14}$$

   where $\hat{f}_k(\xi_1,\xi_2)$ represents all the samples in tassel V_i and N_i is their number.

6. Steps 2-5 or 3-5 are repeated, where the choice on whether the generation of the samples is repeated or not is left to the user (regeneration yields slower convergence and more robust results). The iterations stop when convergence is met according to a selected criterion or a fixed number of iterations is reached. In this paper, the chosen convergence criterion relies on the definition of "Distortion":

$$\Delta\left(\{V_i, f_i\}_{i=1}^N\right) = \sum_{i=1}^{N}\int_{V_i} \|F(\omega) - f_i\|^2_{L^2(\Xi)}\, \mathrm{d}\mathbb{P}_F \tag{3.15}$$

   In particular, convergence is met when the distortion has changed less than 5 percent over the last 5 iterations. Equation (4.3) can be seen as representative of the approximation error, and from a numerical point of view it is computed as follows:

$$\Delta = \sum_{i=1}^{N}\sum_{k=1}^{N_i}\frac{1}{N_i}\|\hat{f}_k(\xi_1,\xi_2) - f_i(\xi_1,\xi_2)\|^2_{L^2(\Xi)} \quad \text{with } \hat{f}_k \in V_i \tag{3.16}$$

7. The probability masses $\mathbb{P}(\Omega_i)$ that correspond to each quantum f_i are determined

after the termination of the iterative scheme and defined as follows:

$$\mathbb{P}(\Omega_i) = p_i = \frac{N_i}{N_{sim}} \tag{3.17}$$

8. Additional iterations with higher values of the parameter N_{sim} may be performed to improve the accuracy on the probability masses $\mathbb{P}(\Omega_i)$.

FQ-IDCVT has effectivelly been utilized for the case of the uni-dimensional Gaussian and non-Gaussian processes, showing the versatility of the approach (Miranda and Bocchini, 2013; Bocchini et al., 2014). The method was designed in a very general way since its inception. Thus the extension to multi-dimensional random functions has been accomplished with relative small changes in the algorithm. The first modification pertains to the first three modules shown in Figure 3.3. In fact the generated realizations have to consider the n-th dimensionality of the domain of the sample functions. The most important adjustment is located in the module where the distance of realization j from all quanta $\{f_i\}_{i=1}^{N}$ is computed. The distance in the $L^2(\Xi)$ norm sense has to be evaluated in all n dimensions.

Finally, the methodology provides an optimal set of sample fields and associated relative weights that can be used for a simulation-based probabilistic analysis with an n-dimensional function.

3.4 Differences between FQ-IDCVT and SROM

As already mentioned, FQ is not the only technique that aims at optimally approximating random functions with a moderate number of samples. In particular, SROM (Grigoriu, 2009) is more popular in the field of civil engineering. For this reason it is important to point out the similarities and differences between the two methodologies.

There are three main conceptual contributions that can be identified in this line of research.

1. The idea of approximating a random function with a "simple function" instead of a parametric representation, as indicated in Equation (3.5). This idea is common to FQ, SROM, and FQ-IDCVT.

2. The formulation of the problem and the choice of the optimality criterion. This is the main difference between FQ and SROM. FQ-IDCVT uses the optimality criterion introduced by FQ in 2002 (mean square optimality), whereas SROM minimizes the discrepancy between selected statistics of the model and the random function being approximated.

3. The way in which the problem is actually solved and the quantizers are computed. This is the main novelty introduced by FQ-IDCVT: the optimization problem is solved in a way that is completely different from what was done for SROM and for previous FQ techniques, such as quantizer design (Luschgy et al., 2010).

Since FQ-IDCVT shares items 1 and 2 with FQ, and only item 1 with SROM, it is presented as a variant of FQ, rather than as a variant of SROM. Depending on the specific application, different techniques may turn out to be the most appropriate. For instance, SROM are particularly suitable for problems involving stochastic partial differential equations, whereas FQ-IDCVT is more appropriate for problems where a mean square approximation is sought, such as those involving hazard analysis.

3.5 Numerical Applications

The applicability to two-dimensional random functions is demonstrated through studying a 2D lognormal field and the classical problem of a panel in traction with random modulus

of elasticity. In the first numerical example the effectiveness of the FQ-IDCVT technique is investigated for different values of the stochastic parameters, in an effort to identify the limits of applicability in terms of the correlation length. In the second example the FQ-IDCVT method is evaluated focusing on the characterization of the resulting output quantities.

3.5.1 2D Lognormal field by SRM

The first example involves a stationary two-dimensional field with non-Guassian marginal distribution. In this example, the technique used for the simulation of the two-dimensional non-Gaussian field generates samples that match a prescribed non-Gaussian marginal distribution (PDF) and a prescribed Spectral Density Function (SDF) based on translation field theory Grigoriu (1995). The methodology was first presented by Shields et al. Shields et al. (2011) and extended later to the multi-dimensional case by the authors Christou and Bocchini (2014a).

The target Spectral Density Function (SDF) is selected as:

$$S_{FF}(\kappa_1,\kappa_2) = \sigma^2 \frac{b_1 b_2}{4\pi} \exp\left[-\left(\frac{b_1\kappa_1}{2}\right)^2 - \left(\frac{b_2\kappa_2}{2}\right)^2\right] \qquad -\infty < \kappa_1 < \infty \quad \text{and} \quad -\infty < \kappa_2 < \infty \tag{3.18}$$

where b_1 and b_2 are parameters proportional to the correlation length of the stochastic field along the axes ξ_1 and ξ_2. Figure 3.4 represents the SDF of a quadrant 2D-1V homogeneous stochastic field when $b_1 = b_2 = 1$m and upper cutoff frequencies $\kappa_{u1} = \kappa_{u2} = 5$ rad/m. The spatial domain of the random function in this numerical application is $[0,40]\text{m} \times [0,40]\text{m}$.

The target marginal distribution is described by the following shifted Lognormal distri-

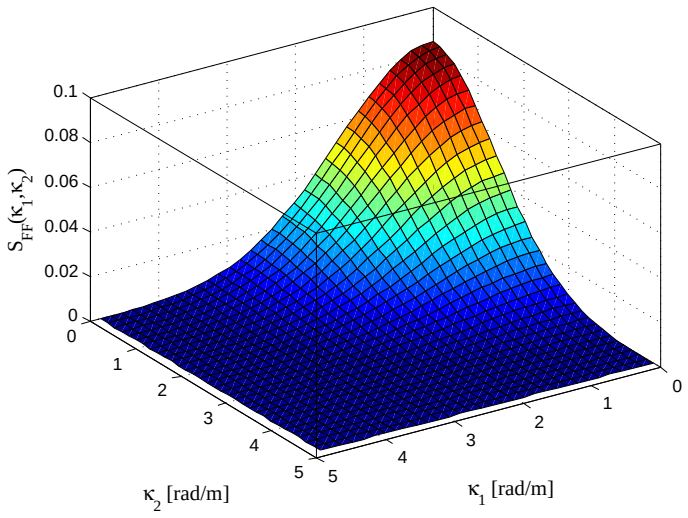


Figure 3.4: SDF of the simulated random field with parameters $b_1 = b_2 = 1$ [m].

bution:

$$\mathrm{PDF}(x) = \frac{1}{\sqrt{2\pi}\sigma_G \bar{x}} \exp\left[-\frac{(\ln \bar{x} - \mu_G)^2}{2\sigma_G^2}\right] \tag{3.19}$$

where:

$$\sigma_G^2 = \ln\left(1 + \frac{\sigma}{\bar{\mu}^2}\right); \qquad \mu_G = \ln \bar{\mu} - \frac{\sigma_G^2}{2}; \qquad \bar{x} = x - \bar{\mu}. \tag{3.20}$$

The values for the parameters $\bar{\mu}$ and σ are chosen as:

$$\bar{\mu} = 1.8; \qquad \sigma^2 = 1 \tag{3.21}$$

For the generation of sample functions of $F(\xi_1, \xi_2)$, the upper cut off frequency is determined according to the following criterion Shinozuka and Deodatis (1996):

$$\varepsilon = 1 - \frac{\int_0^{\kappa_{u1}} \int_0^{\kappa_{u2}} S_{FF}(\kappa_1, \kappa_2)\, \mathrm{d}\kappa_1\, \mathrm{d}\kappa_2}{\int_0^{\infty} \int_0^{\infty} S_{FF}(\kappa_1, \kappa_2)\, \mathrm{d}\kappa_1\, \mathrm{d}\kappa_2} < 0.0001 \tag{3.22}$$

Having specified the probabilistic characteristics of the random field, FQ-IDCVT is

applied considering four cases with different quantizer sizes (i.e. deterministic runs that can be performed) $N = 50, 100$ and 200. The value of the computational parameter N_{sim} is set to $100 \cdot N$ and the algorithm is iterated until the percent of the difference between the average change of the distortion among five succesive iterations is below 5%. For this analysis, the set of samples is kept the same until convergence is met. After convergence is satisfied, an additional iteration is performed with $N_{sim} = 1000 \cdot N$.

An important characteristic of the versatility of the technique is that the parameter N_{sim} does not need to be constant throughout the iterations. However, for higher values of N_{sim} the computation can become very time-consuming. Therefore, for two-dimensional fields it is more efficient to keep the value of N_{sim} small until reaching a near-optimal level of distortion and thereafter refine the results by performing an additional iteration with higher value of N_{sim}.

Convergence plots of the distortion Δ for different values of the quantizer size N and with fixed parameters $b_1 = b_2 = 20$m are provided in Figure 3.5. It is evident that with the convergence parameter being 5%, convergence is met before the twentieth iteration for most simulations. Other results that are not presented herein show that the convergence rate is not particularly affected by the initial seed and when different values of N_{sim} are considered within the range $N \cdot k$, where $k \in [100, 5000]$. However, higher values of the N_{sim} parameter yield a more accurate estimate of the probability masses p_i.

The minimization of the distortion imposed by FQ-IDCVT leads to a holistic convergence of the approximation to the random field, without focusing on a specific moment or probabilistic characteristic (as opposed to other approximate techniques). Nevertheless, a qualitative validation of the FQ-IDCVT technique can be done through the comparison of the probabilistic characteristics of the quantizer and the random function. Such comparison will be presented in the next subsection.

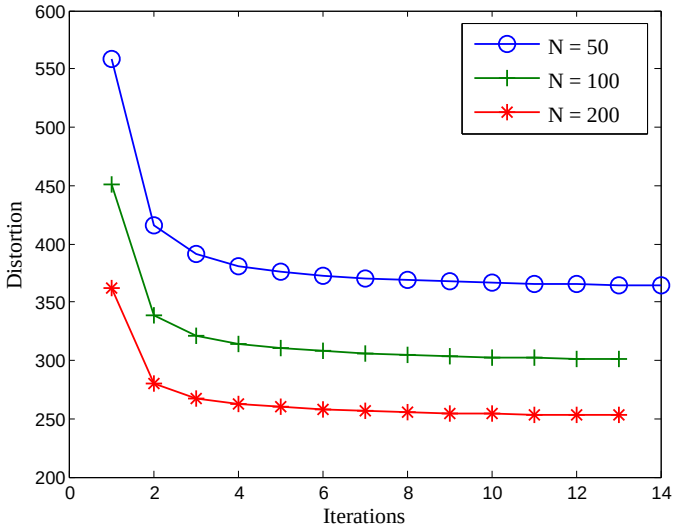


Figure 3.5: Convergence of the Distortion for different values of N and fixed value of parameter $b = 20$ [m].

3.5.1.1 Sensitivity analysis

Every reduced order method is accompanied by limits of applicability, beyond which the approximation error is unacceptable for engineering applications. Similarly, also for FQ-IDCVT it is important to identify the cases where it accurately approximates the investigated random field and those where it should be used with more discretion.

FQ-IDCVT has been successfully applied, on random functions with relatively long correlation length Miranda and Bocchini (2015b); Bocchini et al. (2014). In fact, the repeated use of Equation (3.12) at each iteration yields quanta that are obtained through a series of averages. If the value of N is too small, these averages are performed over a region that is too broad and contains very diverse samples. This issue is an effect of the curse of dimensionality. The random function F is infinite-dimensional because it can take a different value at each of the infinite points of the spatial domain. However, the correlation mitigates this effect. To explain this, two extreme cases can be considered. If the correlation function is zero everywhere (except at the origin), then the value of F at each point

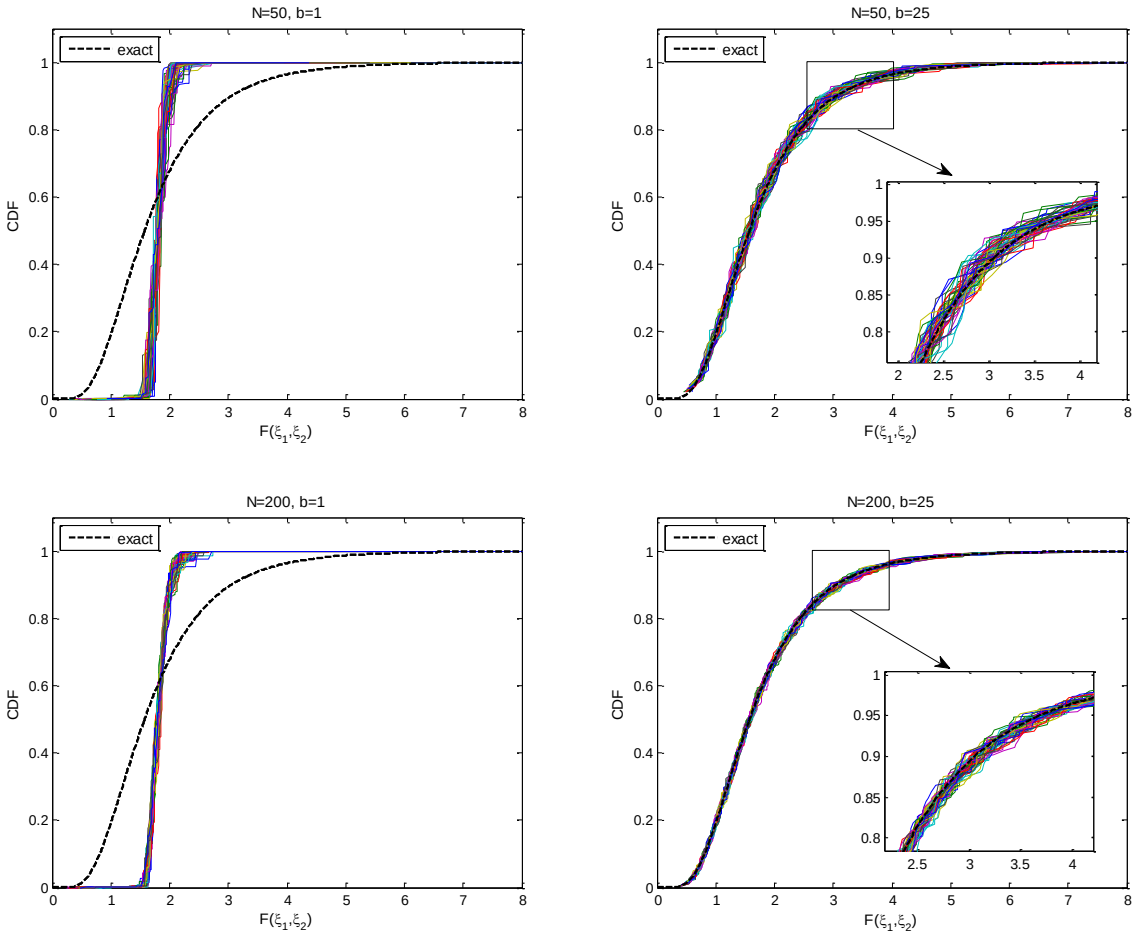


Figure 3.6: Cumulative distribution functions obtained for N=50, $b_1 = b_2 = 1$ m (top left), $N = 50$, $b_1 = b_2 = 25$ m (top right), $N = 200$, $b_1 = b_2 = 1$ m (bottom left) and $N = 200$, $b_1 = b_2 = 25$ m (bottom right). The dashed bold line represents the CDF of the shifted lognormal distribution (exact) and the colored continuous lines represent the CDFs computed at the central grid point of the spatial domain [0, 40] m $\times$ [0, 40] m with 50 different seeds. For small values of the correlation distance the CDF is highly diverged from the exact. In those cases a much higher number of the quantizer size N has to be considered. For higher values of the correlation distance (where the sample is relatively smooth) the outcome of the FQ-IDCVT gives a very good approximation of the exact solution.

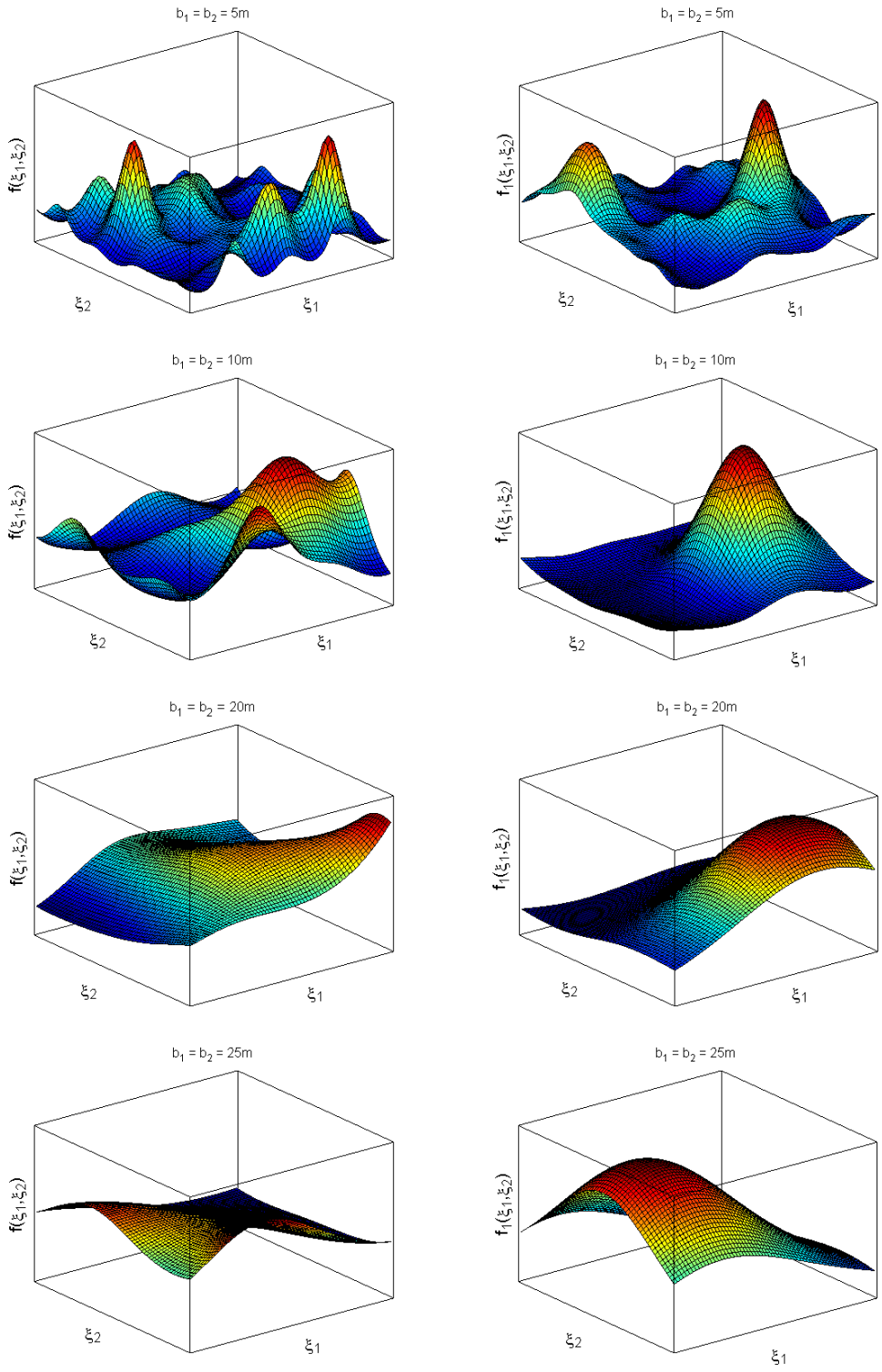


Figure 3.7: Realization and quantizer for $N = 100$ and different values of the correlation lengths b_1 and b_2. The column on the left represents realizations obtained for $b_1 = b_2 = 5, 10, 20$ and 25 m from top to bottom. The column on the right represents the corresponding quantizers obtained for $b_1 = b_2 = 5, 10, 20$ and 25 m from top to bottom.

is an independent random variable. Hence, to completely describe one of its realizations, an array with infinite entries would be required. Instead, if the autocorrelation function is equal to σ^2 everywhere, then all the points are perfectly correlated, all realizations are constant and they can be described by a single value. In this extreme case, one parameter can describe the entire function, so the dimensionality of F would effectively be 1. Similar considerations can be done on the basis of the relevant terms of the Karhunen-Loéve expansion.

FQ approximates F with a finite number of quanta, therefore, with a finite and small number of parameters. Thus, the accuracy of the approximation depends on the difference between the quantizer size and the effective dimensionality of F. The former is described by parameter N, the latter is tuned by the correlation length of the field. A good accuracy is achieved for large values of N or for large values of the correlation length (i.e., for large b_1 and b_2 in this example). Therefore, a parametric analysis has been performed to quantify the accuracy of the FQ approximation of a two-dimensional random function with different values of the quantizer size N and the correlation length.

The Cumulative Distribution Function (CDF) of the quantized model is used for the performance assessment of the technique. CDF's are determined for values of $N = 50, 100$ and 200 and values of $b_1 = b_2 = 1, 5, 10, 15, 20$ and 25 m. It should be noted that, by construction, the quantizer is non-stationary even if the represented random field is stationary. Figure 3.6 compares the approximate CDF's computed for $N = 50$ and 200, $b_1 = b_2 = 1$ m and 25 m at the midpoint of the field. The continuous lines represent the approximate CDF's are computed from FQ-IDCVT with 50 pseudo-random number generator seeds. It is evident that in cases where low values of b_1 and b_2 are considered, FQ-IDCVT yields a poor approximation of the exact marginal distribution. Instead, for higher values of the parameter $b_1 = b_2 = 25$ m, the comparison shows a reasonable level of accuracy even at the tails of the CDF.

Figure 3.7 shows one realization and one representative quantum f_i as computed by FQ for $N = 100$ and different values of the correlation length $b = 5, 15, 20$ and 25. It is observed that the quantum f_i is generally smoother than the realizations, particularly for the cases with smaller correlation length.

For a more rigorous evaluation of the convergence between the approximated and the exact CDF, the Kolmogorov-Smirnov index D_{KS} is utilized Benjamin and Cornell (2014). This can be seen as a check on the uniform convergence (in the analytical sense) of the two CDF's. In particular, D_{KS} measures the maximum discrepancy between the two CDF's and is considered the strictest possible type of convergence:

$$D_{KS} = \max |CDF_{approximate} - CDF_{exact}| \tag{3.23}$$

Figure 3.8 shows the D_{KS} index obtained for different values of N, b_1 and b_2. It is observed that for values of the parameter $b_1 = b_2 > 15$, (b_1 and b_2 are proportional to the correlation length) the FQ-IDCVT technique converges faster and can effectively be utilized even for small values of the quantizer size N.

3.5.1.2 Computational challenges in the extension to 2D

The FQ-IDCVT technique can support any simulation code for the generation of the random samples. However, in this paper the Spectral Representation method is utilized due to its popularity and computational efficiency when the FFT technique is incorporated. Nevertheless, in such a case aliasing has to be avoided, and this can be accomplished when the following criterion is satisfied:

$$M_1 = A_1 \cdot L_1 \quad \wedge \quad M_2 = A_2 \cdot L_2 \quad \wedge \quad A_1, A_2 \geq 2 \tag{3.24}$$

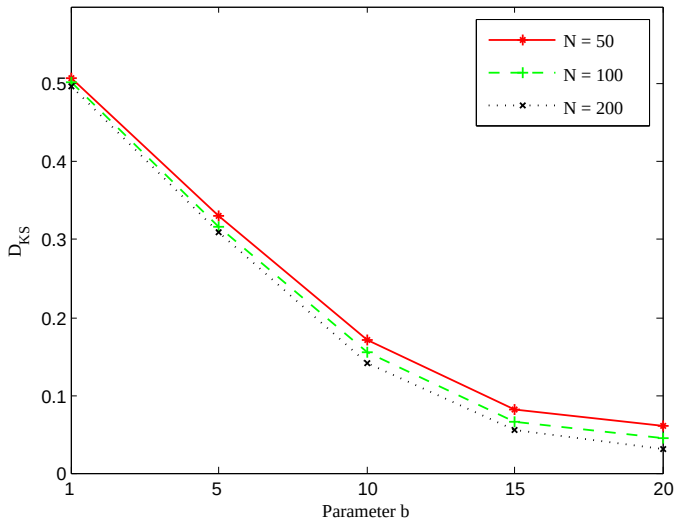


Figure 3.8: Kolmogorov-Smirnov index obtained for different quantizer sizes N, and different values of the correlation length b_1 and b_2. D_{KS} measures the uniform convergence (in the analytical sense) of the approximate CDF to the exact. FQ has better accuracy when the random functions have long correlation length.

where M_1 and M_2 are the number of points of the simulated stochastic field discretization along the ξ_1 and ξ_2 axes, respectively, while L_1 and L_2 are the number of discretization points in the wavenumber domain along axes κ_1 and κ_2, respectively.

The first challenge to be addressed is the proper adjustment of the cutoff frequencies κ_{u1} and κ_{u2} and the number of points along ξ_1 and ξ_2 axes (i.e. M_1 and M_2), when samples with high correlation lengths are considered. The selection of these parameters has to be such, that the discretization of the space domain of interest is the same for all the simulated stochastic fields with different correlation lengths. Small values of the cutoff frequency determine the generation of very large samples when the FFT technique is utilized. Then, the samples have to be cropped and still have a sufficient number of points within the cropped region to be comparable with the samples characterized by small correlation length (i.e smallest period when FFT is utilized). For consistency, all the simulated stochastic fields are considered to have the same discretization of their Spectral Density (i.e., fixed values of L_1 and L_2 which for this application are both consider to be equal to 32).

From Equation (3.18), for large values of the parameters b_1 and b_2 the area under the SDF concentrates at low values of the wave numbers κ_1 and κ_2. Among the values of cutoffs that satisfy Equation (3.22), the smallest one should be selected, otherwise the discretization in the relevant part of the SDF would be very coarse for large values of b_1 and b_2. As a result, the poor discretization of the SDF would yield random samples with probabilistic characteristics different from the desired marginal distribution. The values of κ_{u1} and κ_{u2} for different sets of parameters b_1 and b_2 are depicted in Table 3.1.

The stochastic field samples generated by SRM have period:

$$\begin{aligned} T_{\xi_1 0} &= \frac{2\pi}{\Delta\kappa_1} \quad \text{along } \xi_1\text{axis} \\ T_{\xi_2 0} &= \frac{2\pi}{\Delta\kappa_2} \quad \text{along } \xi_2\text{axis} \end{aligned} \tag{3.25}$$

and $\Delta\xi_1$, $\Delta\xi_2$ and $\Delta\kappa_1$, $\Delta\kappa_2$ are related in the following way:

$$\Delta\xi_1\Delta\kappa_1 = \frac{2\pi}{M_1} \quad \text{and} \quad \Delta\xi_2\Delta\kappa_2 = \frac{2\pi}{M_2} \tag{3.26}$$

The wavenumber increments $\Delta\kappa_1$, $\Delta\kappa_2$ are defined as:

$$\Delta\kappa_1 = \frac{2\pi}{L_1} \quad \text{and} \quad \Delta\kappa_2 = \frac{2\pi}{L_2} \tag{3.27}$$

Combining Equations (3.27), (3.26) and solving for ${}^{M_1}/_{L_1}$, the parameter A_1 can be determined. Similar results can be obtained for parameter A_2 when solving for ${}^{M_2}/_{L_2}$:

$$A_1 = \frac{2\pi}{\kappa_{u_1}\Delta\xi_1} \quad \text{and} \quad A_2 = \frac{2\pi}{\kappa_{u_2}\Delta\xi_2} \tag{3.28}$$

In this example the random function with $b_1 = b_2 = 1$ m yields realizations with period of 40m in each direction. The space increments $\Delta\xi_1$ and $\Delta\xi_2$ computed for this case are

used in Equation (3.28) for the computation of parameters A_1 and A_2. And again, the goal is to produce realizations for different values of b which have the same number of points within the domain $[0,40;0,40]$m. This is accomplished by increasing M_1 and M_2 so that the domain $[0,40;0,40]$m is a grid of $(2\cdot 32)\times(2\cdot 32) = 4096$ points. Note that for large correlation lengths, this yields big samples, which significantly slow down the computations. As an example, the values of the discussed parameters for the different cases of the correlation length are provided in Table 3.1.

Table 3.1: Input parameters for the sample generation of the two-dimensional random field via SRM considering different correlation lengths.

$b_1 = b_2$	$M_1 = M_2$	$\kappa_{u1} = \kappa_{u2}$
1	64	5.0
5	320	1.0
10	640	0.5
15	1088	0.3
20	1280	0.25
25	1600	0.2

3.5.2 Two-dimensional panel with random mechanical properties under plane stress

The second example regards a square panel with uncertain Young modulus under plane stress. The panel is loaded with a uniform load at one edge and constrained with pinned connections at the opposite edge, as depicted in Figure 3.9. The example is based on the application presented in Falsone and Impollonia (2002). In specific, the panel has side dimensions $\ell = 1$m and thickiness $t = 0.1$m. For the linear static analysis of the structure, a mesh of 100 four-node quadrilateral finite elements with 8 degrees of freedom (horizontal and vertical displacement at each node) is considered. The Poisson ratio is assumed equal to 0.3. The Young modulus is considered uncertain and is modeled as a two-dimensional

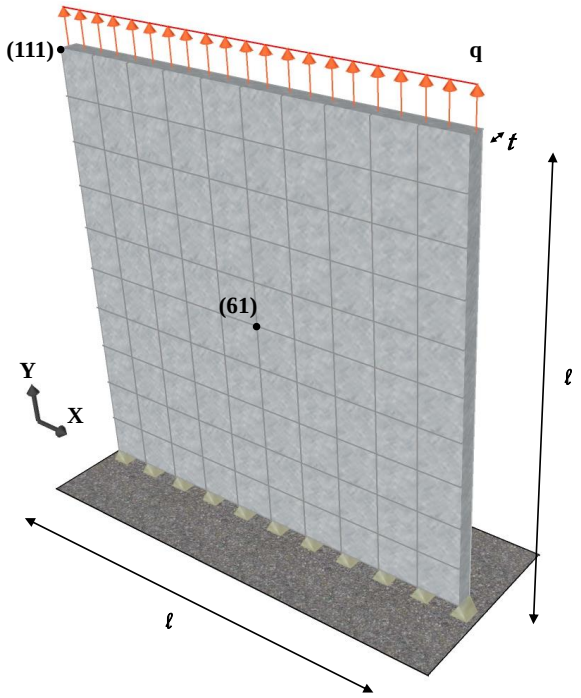


Figure 3.9: Two-dimensional panel structure with uncertain Youngs modulus.

stochastic field with constant mean value $\bar{E} = 200 \cdot 10^9$ Pa :

$$E(x,y) = \bar{E}\,(1 + \alpha(x,y)) \tag{3.29}$$

where $\alpha(x,y)$ is a Gaussian field which has zero mean and covariance given by:

$$\Sigma_\alpha\,(|\boldsymbol{\Delta x}|) \equiv \sigma^2 \rho\,(|\boldsymbol{\Delta x}|) = \sigma^2 \exp\left(-\frac{|\boldsymbol{\Delta x}|}{\lambda}\right)^2 \tag{3.30}$$

where $\boldsymbol{\Delta x}$ is the distance between two points of the field, σ^2 is the variance and $\lambda = 1.0\ell$ is the correlation length. The discretization of the random field is made according to the midpoint method, which yields stepwise realization of the field with discontinuities along the element boundaries and distances measured at the centroid of every finite element.

Each sample of the probabilistic Young modulus is associated with one deterministic analysis. In order to more effectively run the large number of deterministic analyses, while

better controlling the input and assessing the output quantities, an in-house finite element code was developed. The accuracy of this finite element code was verified and validated against the commercial software Abaqus Version (2013).

Having determined the probabilistic characteristics of the random field, the realizations are generated using FQ-IDCVT and MCS. In the case where FQ-IDCVT is used, the output of the deterministic analysis are weighted according to the probability mass associated with the quantum used as input. In this example, FQ-IDCVT is applied considering four cases with different quantizer sizes $N = 50, 100, 200$ and 500. The value of the computational parameter N_{sim} is set to $100 \cdot N$ and the algorithm is iterated until the rate of convergence (as explained in the previous example) is lower than 2.0%. In this example the samples are generated once at the onset of the procedure and the time when convergence is met, an additional $N_{sim} = 1000 \cdot N$ samples are generated to improve the accuracy of the probability masses associated with each quanta.

Looking at the quantized input quantities, the approximation of the Young modulus of a corner finite element is not as accurate compared to the case where the center finite element is considered. In general, when FQ-IDCVT is utilized for the generation of the input fields, the center regions of the resulting quanta always exhibit higher accuracy compared to the peripheral regions of the field. Because of the correlation, each point of the grid that discretizes the field tends to have values that are similar to those of the surrounding grid points. For peripheral points, part of the neighbor grid points are actually out of the domain. Hence, they tend to influence the clustering phase less than the central points (which can count also on the weight of all the neighbor points with similar values). As a result, central points really drive the clustering phase and thus the quanta tend to represent better these points. However, this observation applies only to the input quantities, whereas the accuracy of the output quantities (i.e., vertical displacement) is mostly affected by the mechanical filter itself. In fact, Figure 3.10 compares approximate CDFs computed by FQ-IDCVT for

$N = 100$ with the exact solution. More specifically, the corner node (i.e., 111) and center node (i.e., 61) of the plate are evaluated using the marginal distribution function. For the other nodes of the plate, the CDFs of the vertical displacements exhibit similar accuracy.

The second validation of the methodology is based on the evaluation of the autocorrelation of the vertical displacements among the nodes of the plate. A comparison was made between the correlation function yielded from FQ-IDCVT and the exact solution. The autocorrelation of a two-dimensional field is a four-dimensional function $R(\xi_{11}, \xi_{21}, \xi_{12}, \xi_{22})$ which cannot be easily depicted. In the later function, the first subscript refers to the field axes in the spatial domain and the second subscript refers to the point of the field. Two simplified representations are considered: (1) the autocorrelation $R(\bar{\xi}_{11}, \xi_{21}, \bar{\xi}_{12}, \xi_{22})$ among single rows (or columns) of grid points, treated as 1D fields; (2) the autocorrelation between a fixed point and all the other points of the grid $R(\bar{\xi}_{11}, \bar{\xi}_{21}, \xi_{12}, \xi_{22})$. Both approaches yield good approximations of the exact autocorrelation function. Figure 3.11 represents the comparison between the quantized and the exact autocorrelation functions of the corner vertical displacement with the vertical displacements of all the other nodes of the structure. The approximation of the exact autocorrelation is sufficiently accurate even for a small quantizer size (i.e., $N = 100$).

Figure 3.12 shows Box and Whisker plots of the autocorrelation on the vertical displacement of the corner node (i.e., 111) with the center node (i.e., 61). The results from FQ-IDCVT technique are obtained for $N = 50, 100, 200$ and 500 and compared with the outcomes obtained by MCS for $N = 50, 100, 200, 500, 1.000, 5.000$ and 10.000. For each case the analysis is repeated 50 times using different seeds. Bias error (BE) and stochastic error (SE) are popular metrics to assess the accuracy and robustness of a statistical estimator. In this case they are used to quantify the quality of the representation of the autocorrelation. In particular, they are applied to the autocorrelation between the value of the displacement at the corner and center nodes. Similar results have been obtained for other

pairs of nodes. BE is the average difference between the various instances of the estimator (obtained using the 50 different seeds) and the exact value, which in this case has been computed by MCS with 500,000 samples because a closed form solution is not available:

$$\mathrm{BE}_{R_{vv}(v_{111},v_{61})} = \frac{1}{\mathrm{nseeds}} \sum_{iseed=1}^{\mathrm{nseeds}} R_{vv}^{iseed}(v_{111}, v_{61}) - \bar{\mu}_{R_{vv}} \tag{3.31}$$

where: nseeds is the total number of estimations, $R_{vv}(v_{111}, v_{61})$ is the autocorrelation between the value of the displacement at the corner and center node and $\bar{\mu}_{R_{vv}}$ is the exact mean value of $R_{vv}(v_{111}, v_{61})$ computed by MCS with 500,000 samples. MCS is known to be unbiased, by construction. Instead, the FQ-IDCVT technique produces slightly biased results. However, the absolute value of BE is very small, even for small quantizer sizes. In this example, it always lies within the range $[2.5-5]\cdot 10^{-8}$ compared to values of the autocorrelation in the order of $5\cdot 10^{-7}$. A bias error smaller than 10% on a second order estimator like the autocorrelation with only 50 samples for a 2D field is perfectly satisfactory. The value of BE drops if larger quantizer sizes are used. In terms of SE, FQ-IDCVT performs significantly better compared to MCS. SE is computed as the standard deviation of the estimator (i.e., the autocorrelation between corner and node displacement in this case) obtained for different estimations (i.e., for 50 different seeds in this case):

$$\mathrm{SE}_{R_{vv}(v_{111},v_{61})} = \sqrt{\frac{1}{\mathrm{nseeds}} \sum_{iseed=1}^{\mathrm{nseeds}} [R_{vv}(v_{111}, v_{61})_{iseed} - \mu_{R_{vv}}]^2} \tag{3.32}$$

where: $\mu_{R_{vv}}$ is the mean value of $R_{vv}(v_{111}, v_{61})$ between the estimates. Looking at the sizes of the boxes in Figure 3.12, it appears that FQ-IDCVT with quantizer size N = 50 yields the same SE of MCS with 5,000 samples. The results in Figure 3.13 shows the trend of SE and clearly quantifies the superior performance of FQ-IDCVT.

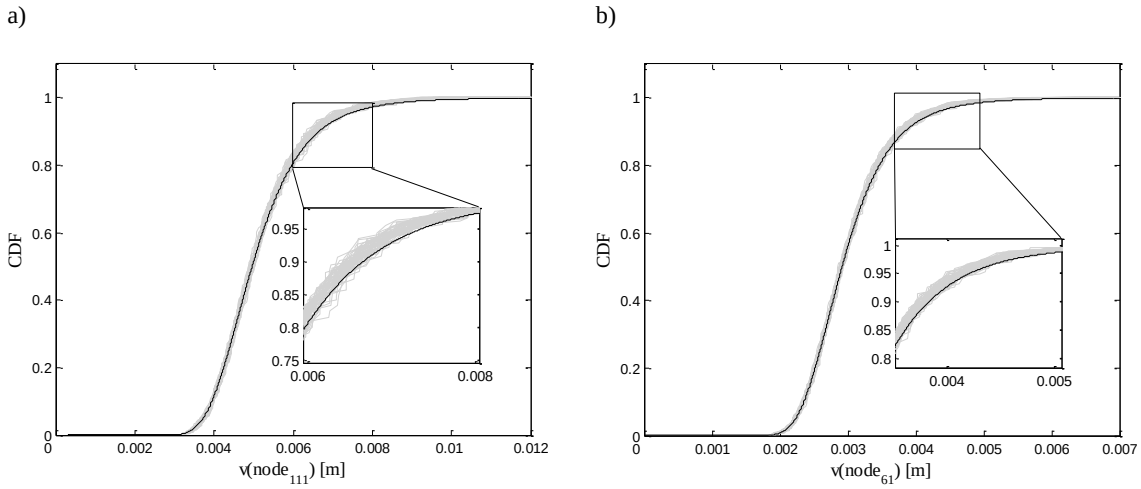


Figure 3.10: Cumulative distribution functions obtained for N=100. The figure on the left is the CDF of the corner node (i.e., 111) and the figure on the right is the CDF of the center node (i.e., 61). In both figures the continuous black line represent the CDF obtained from 500,000 samples (exact) whereas the gray continuous lines represent the CDF's computed from FQ with sample size N=100 and 50 different seeds. For this example where the correlation distance is equal to the the size of the plate, the outcome of the stochastic static analysis, considering a small quantizer size (N=100), gives a very good approximation of the exact solution.

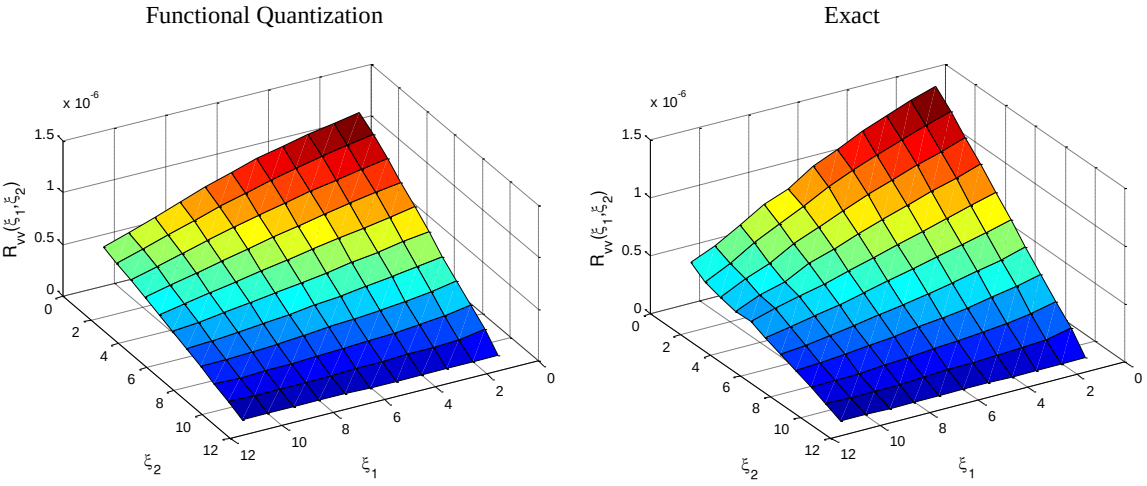


Figure 3.11: Comparison between the optimal quantization (left) and the exact (right) autocorrelation function for quantizer size N=100 of the corner node (i.e., 111) vertical displacement with the vertical displacements of all the other nodes of the structure. When FQ-IDCVT is used, the autocorrelation is always underestimated, however, in this case the approximation is very good for a relatively small number of quantizer size.

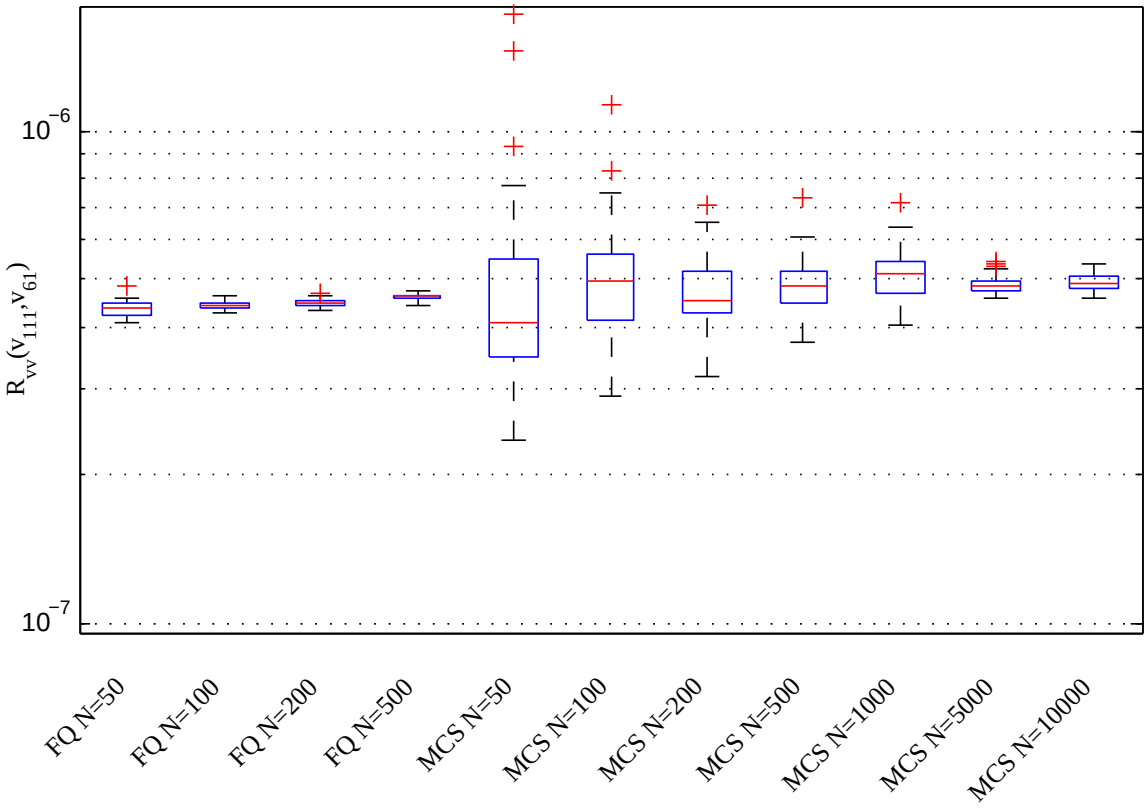


Figure 3.12: Box and Whisker plots of the autocorrelation on the vertical displacement between the corner and the center node. For FQ-IDCVT the results are obtained for $N = 50, 100, 200$ and 500 and compared with the outcomes obtained by MCS for $N = 50, 100, 200, 500, 1.000,$ and 5.000. For small values of the quantizer size, the stochastic error on the autocorrelation is significantly smaller. For example, MCS with 10,000 samples provides similar stochastic error with FQ-IDCVT and quantizer size $N = 50$

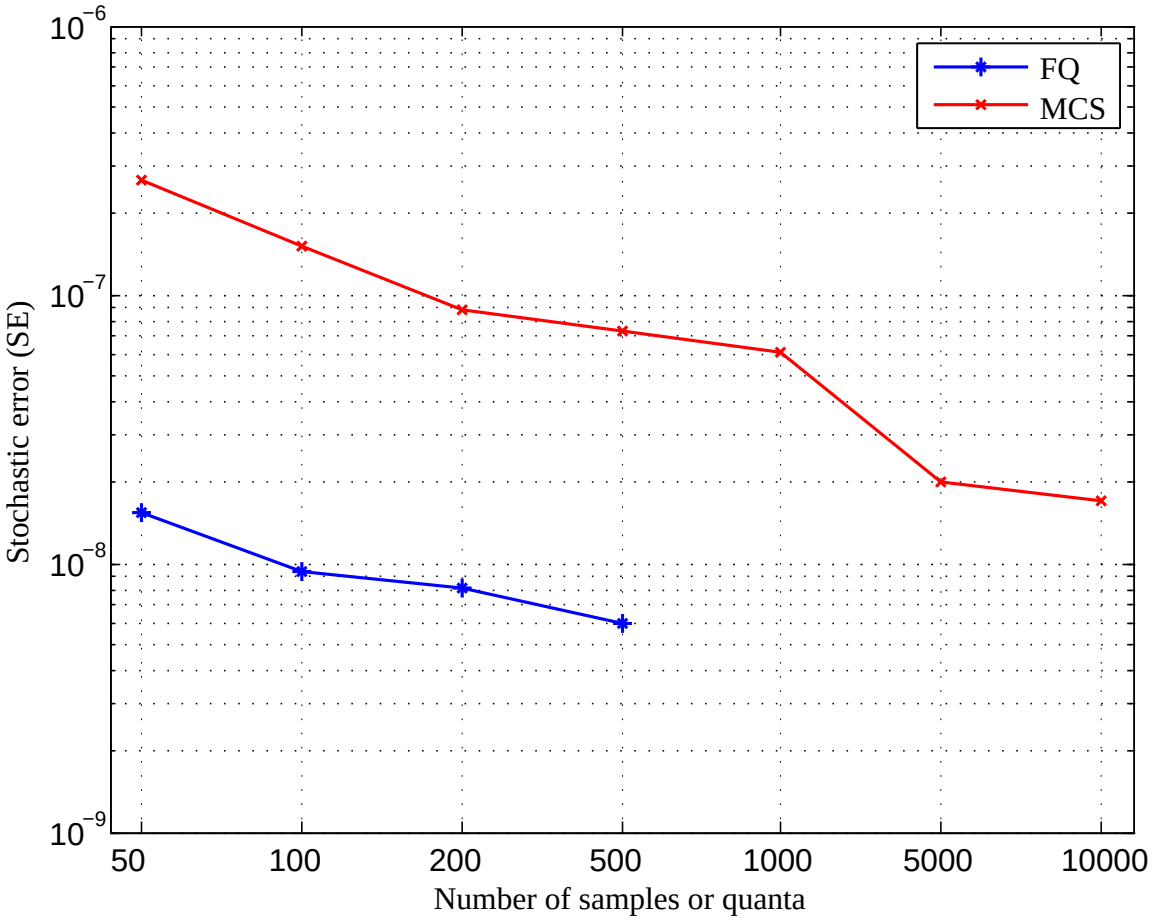


Figure 3.13: Stochastic error of the autocorrelation on the vertical displacement between the corner and the center node. For FQ-IDCVT the results are obtained for $N = 50, 100, 200,$ and 500, with each quantizer size being evaluated fifty independent times. These results are compared with the stochastic error obtained by MCS for $N = 50, 100, 200, 500, 1.000, 5.000$ and $10,000$. The different set of sample sizes have been run for fifty independent times using different seeds each time. In terms of the percent error of the same metric, MCS with 1000 and 10000 samples exhibit 12.56% and 3.51% error respectively, whereas FQ with 50 quanta exhibits 3.15% error.

3.6 Concluding Remarks

A novel methodology has been presented for the functional quantization of multi-dimensional random functions. FQ-IDCVT has shown its versatility computing quantizers of uni-dimensional Gaussian and non-Gaussian, stationary and non-stationary random processes in the past. In this paper, the extension to the multi-dimensional space is explored and two applications (a two-dimensional non-Gaussian random field and a two-dimensional panel structure with uncertain Young modulus under plane stress state) have been presented to demonstrate its applicability for engineering applications and investigate its accuracy. FQ was established as a very general approach and thus some moderate adjustments had to be made in the algorithm to consider the multi-dimensionality as explained in Section 3.3. To the best of the authors' knowledge, none of the techniques available in the literature has ever been implemented in a way to provide optimal quantizers for non-Gaussian random functions defined over multi-dimensional spaces.

A quantitative investigation on the accuracy of the FQ-IDCVT technique for different values of the quantizer size N and the correlation length was performed. The method operates particularly well in cases of random functions with relatively large correlation length. In cases with small correlation length the probabilistic characteristics of the quantizer do not accurately approximate those of the desired random function, but they are still the best possible approximation (in the mean square sense) for a fixed number of deterministic analyses that can be run.

For different values of the quantizer size N the convergence rate of the distortion does not change significantly, but for higher values of N_{sim}, the quanta are smoother and the accuracy of the computed weights (Equation 3.17) is increased. However, when two or more dimensional fields are considered, setting $N_{sim} > 100 \cdot N$ can be computational very expensive and thus it is suggested to increase N_{sim} as fine tuning, only after convergence on

the distortion has been met.

In the first application, a detailed description of a few computational challenges was provided when discussing the accuracy analysis, to facilitate the implementation of the proposed methodology. In particular, the determination of the upper cutoff frequency, the discretizations in the space and wave number domains while the FFT technique is utilized in the SRM method and the difficulties associated with estimating the non-stationary CDF in the investigated domain were presented.

Compared to naive MCS, the proposed technique allows to reduce by some orders of magnitude the number of required samples and associated deterministic runs, for a similar quality of the probabilistic characterization of the results. Compared to other techniques for the representation of random functions with few samples, FQ has the characteristic of providing optimal approximations in the mean square sense, which is important for several applications, such as risk analysis. Compared to other FQ-based techniques, FQ-IDCVD is the only one applied also to non-Gaussian and multi-dimensional functions, in addition to the ability to handle non-stationarity/homogeneity. All this is achieved with a very simple computational procedure presented in Section 3.3, which can be easily paired with any subroutine for the generation of random samples.

Chapter 4

Effective sampling of spatially correlated intensity maps using Hazard Quantization: application to seismic events

4.1 Introductory remarks

In the past years, there has been increased attention on regional hazard analysis as a response to the outcomes of severe natural catastrophes that occurred in the recent past. Because of the intensity and interdependency of the long-reaching effects of these catastrophes at the regional scale, researchers realized that community infrastructure needs to be studied comprehensively.

Examples of natural disasters with such effects include the 2011 Tohoku earthquake and tsunami in Japan and the 2012 Hurricane Sandy. Although at both locations the direct losses were destructive, it is the indirect losses that were most pervasive. Tohoku experienced $319 billion in direct losses, and more than $619 billion in indirect losses (Tesfamariam and Goda, 2013) from which it is still recovering. Likewise, after Sandy, indirect economic losses were estimated at about $16.3 billion and business interruption losses were quantified between $10.8 and $15.5 billion (Kunz et al., 2013).

Many structures are exposed to multi-hazard loads during their service life and thus an effective technique for the regional multi-hazard analysis is auspicious. The term multi-hazard refers to hazards that can be correlated or uncorrelated, sequential or non-sequential. An example of a sequential and correlated event affected the Japanese Sendai region's built

environment, which was stricken by the Tohoku earthquake, and immediately thereafter by a powerful tsunami. Conversely, an urbanized region affected by uncorrelated and non-sequential events is the U.S. East Coast's Charleston region, which is prone to both earthquakes and hurricanes. The technique proposed herein aims to be used for uncorrelated and non-sequential hazards. In particular, this paper focuses on seismic hazard applications.

An accurate regional seismic hazard model is required for an accurate seismic risk assessment, which may include infrastructure reliability assessments (Dueñas-Osorio and Vemuru, 2009), risk analyses (Padgett et al., 2010), and community resilience predictions (Karamlou and Bocchini, 2016). The joint earthquake hazard over two or more points of a lifeline is crucial when investigating the simultaneous failure of a certain combination of elements of the network which may result in the failure of the whole network (Rhoades and McVerry, 2001; Bocchini and Frangopol, 2011b). One well-known method for seismic risk evaluation uses the HAZUS-Earthquake tool (Department of Homeland Security, 2003); however, its probabilistic analysis lacks the consideration of correlated seismic excitations, correlated structural capacities and correlated damage costs. For these reasons, and the complexity of the overall analysis, the most popular and appropriate way to address these probabilistic studies is through the use of the simulation-based techniques. These techniques are usually combined with optimal sampling strategies that identify a small stochastically representative catalog of intensity measure (IM) maps, which can be used as an integral part of the traditional Performance-Based Earthquake Engineering (PBEE) framework (McGuire, 2004; Kramer, 1996; McGuire, 2008).

The ground motion correlation describes the similarity of the IM at the various locations of the region and is essential for the estimation of the distribution of losses. The seismic IM maps are usually generated by using a ground motion prediction equation (GMPE), which in common practice is a function of three components. The first component is the median value of the ground motion, which introduces a correlation among sites that are

close, gradually decaying with distance.

The second and third components capture the randomness of the observed amplitudes around the median at different sites. In particular, the second component, called inter-event residuals" or "between-earthquake variability," models the variability of the ground shaking intensities observed for different events with similar magnitudes and rupture mechanisms. This accounts for the fact that even the most sophisticated GMPEs adopt a simplified representation of the earthquake itself. The very complex relative motion of different lithospheric plates (i.e., infinite degrees of freedom) is represented with a finite number of parameters (e.g., location of the epicenter, magnitude, depth, type of rupture mechanism). Therefore, even after these parameters have been fully defined, there is a large amount of uncertainty in the event model, which is represented by the inter-event residuals. The inter-event correlation of earthquake ground motion can be modeled as proposed by Wesson and Perkins (2001).

The third component of a GMPE is called "intra-event residuals" or "within-earthquake variability and it addresses the additional variability of the observed IM amplitudes from site to site within the same seismic event. This variability is mainly due to the incomplete knowledge (and, therefore, incomplete model) of the geological properties of the region. A comprehensive description of the peculiarities of the propagation of seismic waves from the epicenter to the various sites would require complete knowledge of the mechanical properties of every rock and grain of soil in the region (again, infinite parameters). Obviously, in practice the GMPEs describe the seismic effect at each site using a finite number of parameters (e.g., soil type at the site). The uncertainty associated with the parameters that are not directly modeled is captured by the intra-event residuals. When dense observations from numerous earthquakes are available for the region of interest, several methods can be adopted to estimate this within-earthquake correlation of ground motion residuals (Bommer et al., 2003; Wang and Takada, 2005; Goda and Hong, 2008b; Baker and Jayaram,

2008; Jayaram and Baker, 2009). In summary, inter-event residuals account for the incomplete description of the seismic event, and intra-event residuals account for the incomplete description of the site.

The effect of ground motion correlation on the loss assessment of interdependent systems and distributed building assets has been a field of extensive research in past years. Studies that used either comprehensive databases of strong ground motions (Lee and Kiremidjian, 2007; Goda and Hong, 2008a; Goda and Atkinson, 2009) or synthetic earthquake catalogs (Wesson and Perkins, 2001; Bommer and Crowley, 2006; Crowley and Bommer, 2006; Park et al., 2007; Goda and Hong, 2008a, 2009) showed that the ground-motion correlation can significantly affect the probability of joint damage and the characteristics of the loss distribution. The rate of decay of the spatial correlation with separation distance was also investigated for different historical events (Goda and Hong, 2008a; Goda and Atkinson, 2009). Furthermore, the relative influence of the intra-event and inter-event correlations on the estimates of seismic losses for distributed portfolios has been shown to vary with the level of hazard and thus return period (Sokolov and Wenzel, 2011a,b).

The seismic hazard model based on the Cornell-McGuire approach (Cornell, 1968; McGuire, 1976) is a tool commonly used by researchers to assess the seismic risk of spatially distributed building assets. Although this model yields accurate results for individual locations and allows to evaluate the seismic hazard contour maps, it has some disadvantages that can be overcome with the use of simulation-based analysis. Probabilistic simulation naturally accounts for spatially correlated correlated ground motion characteristics and it can easily consider non-homogeneous and non-Poissonian earthquake occurrence models. A few of these simulation-based techniques have been developed for regional risk assessment that try to incorporate the spatial correlation of ground motion, as will be further explained in the following section.

A pliable multi-hazard tool is presented in this paper, which accounts for the spatial correlation through the optimal sampling of IM maps. One of the defining characteristics of the proposed technique is its ability to consider the IM maps of a natural hazard as a random field. This allows the technique to take advantage of methodologies developed within the context of random field theory. An effective methodology of this family called "Functional Quantization" (FQ) is used to provide an optimal (in the mean-square sense) approximation of a random field with a relative small set of samples (Luschgy and Pagès, 2002; Miranda and Bocchini, 2015a; Christou et al., 2016b). The methodology suggested in this paper, named "Hazard Quantization" (HQ) is based on the FQ technique, and this yields several benefits. First, it confers versatility to HQ, in that it can be applied to different hazard types, given the availability of sample IM maps. Second, the representation of the regional hazard is supported by proofs of optimality. Finally, the resulting procedure is elegant and simple to implement. For instance, all the random parameters involved in the hazard simulation module are treated equally, without the need of specialized simulation techniques or any hierarchical sampling, as required by other methodologies in the same field.

4.2 Overview of available methodologies for regional hazard analysis

Several techniques have been developed recently for probabilistic seismic hazard analysis. The suggested methodology is proposed for multi-hazard analysis involving uncorrelated and non-sequential hazard events. A brief description of the selected techniques for regional hazard analysis follows.

One way of categorizing the techniques in contemporary literature is to distinguish between the simulation-based and the non-simulation-based methods, as shown in Figure 4.1

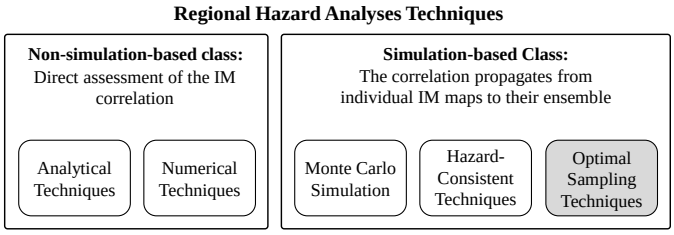


Figure 4.1: Current techniques for regional hazard analysis are categorized in two classes: non-simulation-based and simulation-based. The former class includes two subgroups containing analytical and numerical techniques, and the latter includes three subgroups with Monte Carlo Simulation (MCS), hazard-consistent techniques and techniques based on optimal sampling (Bocchini et al., 2016). The last box is shaded because HQ belongs to that group.

(Bocchini et al., 2016). The non-simulation-based class includes two subgroups: analytical (Wesson and Perkins, 2001; Gardoni et al., 2003) and numerical methodologies (Bocchini and Frangopol, 2011b). Wesson and Perkins (2001), for example, presented a non-simulation method, which directly uses the information of the spatial correlation of seismic intensities for the computation of the distribution of losses for a portfolio of building assets. Likewise, for an accurate analytical risk assessment of a transportation network, Lee and Kiremidjian (2007) considered two types of ground motion correlation models: one that assumes non-distance dependence between the sites of interest and a second which accounts for distance dependence. These models are appealing and simple, yet rely on substantial assumptions. Other researchers utilized numerical techniques based on random field theory to assess bridge damage levels in order to directly control the ground motion and damage correlation (Bocchini and Frangopol, 2011b).

The second and most commonly used family of techniques, the simulation-based class, is based on the idea that spatial correlation is embedded in all real and realistic individual IM maps. Therefore, if these maps are used in a simulation-based fashion, the correlation present in each one of them carries over to the subsequent steps of a PBEE-like framework. Methods based on Monte Carlo simulation (MCS) were first considered in seismic hazard analyses almost two decades ago (Ebel and Kafka, 1999; Musson, 1999, 2000). At the time,

however, the available GMPEs did not consider spatial correlation for the intra- and inter-event residuals. This limitation was investigated extensively in the early 2000s (Rhoades and McVerry, 2001). In 2006, Crowley and Bommer (2006) used MCS for the seismic loss prediction of the North side of the Sea of Marmara in Turkey, while accounting for the spatial variability of the ground motion intensity measures. If a large number of stochastic scenarios is used (large enough to be representative of the hazard in the region), brute-force MCS provides an accurate and unbiased characterization of the regional hazard that can be taken as reference solution. Brute-force MCS, as used by Crowley and Bommer (2006), is conceptually straightforward and tractable, though at the expense of increasing the required computational resources, effort and time of the subsequent analysis phases.

A subgroup of the simulation-based techniques stems from the hazard-consistent probabilistic method that was first introduced by Chang et al. (2000). A small number of earthquake events was selected semi-manually and the annual occurence probability was adjusted so that the probability of exceedance of the reduced set of IM maps matched the probability of exceedance obtained from a comprehensive probabilistic analysis, like the traditional USGS hazards maps or a brute-force MCS. More recently, similar techniques have been developed based on probabilistic optimization (Campbell and Seligson, 2003; Legg et al., 2010; Apivatanagul et al., 2011; Vaziri et al., 2012; Han and Davidson, 2012) that minimizes for all sites of interest the discrepancy between the hazard curve obtained from the weighted set of selected IM maps and the reference hazard curve. In particular Han and Davidson (2012) provided an outstanding review of this entire research field.

Techniques from the hazard-consistent probabilistic subgroup have been applied also to regional hazard analyses for different natural disasters. For instance, hurricane wind hazard in the state of North Carolina was determined in Legg et al. (2010) by using a mixed-integer linear optimization formulation. One year later, this methodology was extended to select hurricanes whose annual occurrence probabilities match both the wind and surge hazards

(Apivatanagul et al., 2011). In Vaziri et al. (2012) the authors applied the hazard-consistent probabilistic approach to identify earthquake scenarios for Tehran. Later, a technique was developed based on the methodology presented in Vaziri et al. (2012) to reduce the set of ground motion maps for a specified earthquake event, minimizing once again the errors on the annual probability of exceedance between the reduced set of ground motion maps and the reference MCS solution (Han and Davidson, 2012). The latest development of this class of simulation-based techniques is based on Han and Davidson's four step generation-reduction method, the difference being that Manzour et al. (2016) reduce the number of earthquake scenarios using both reduction methods presented by Han and Davidson (2012).

Another group of simulation-based techniques uses variance reduction methods for an effective sampling of a set of earthquake scenarios. Kiremidjian et al. (2007) introduced importance sampling for the simulation of earthquakes over a range of strong-magnitude events, while Jayaram and Baker (2010) used importance sampling also for magnitude simulation and the ground motion inter- and intra-event residuals. In the same paper, Jayaram and Baker used *k*-means clustering to group similar IM maps and further reduce their number. The HQ technique presented in this paper, which belongs to the same category, shows similarities with the *k*-means clustering technique, but it is based on a different perspective, which will be discussed in the following section.

4.3 HQ methodology

A few studies have mentioned the idea of directly modeling IM maps as random fields. For instance, Wang and Takada (2005) analyzed dense observation data of earthquakes that occurred in Japan and Taiwan and proposed a correlation model for the residuals after verifying that they constitute a homogeneous two-dimensional stochastic field. Along this line , a key characteristic of the proposed HQ methodology that distinguishes it from the

techniques listed in the previous section is that HQ rigorously considers the IM maps as random fields.

Because of the inherent variability of their essential features, natural hazards are considered random phenomena. This randomness makes their characterization and the modeling of their effects over a geographic region a complex problem with patterns of interdependencies. The set of values of an IM of interest resulting from a natural hazard over a geographic region is characterized by a certain degree of uncertainty and correlation, and this is precisely the definition of a 2D random field. Therefore, we can consider as a random field the maximum wind speeds associated with a hurricane throughout a study region, as well as the level of seismic ground shaking. The latter is usually modeled with the popular empirical GMPEs for seismic spectral accelerations, in which the random field is broken down into its (random) median values over the 2D spatial domain and the (random) fluctuations around the median, the already mentioned residuals. Let $M_i(\text{lon},\text{lat})$, $\text{lon} \in [\text{lon}_{min}, \text{lon}_{max}]$, $\text{lat} \in [\text{lat}_{min}, \text{lat}_{max}]$ be the ith realization of a 2D non-homogeneous, non-Gaussian random field representing the intensity measure over a region, where $\text{lon}_{min}, \text{lon}_{max}, \text{lat}_{min}$ and lat_{max} are the geographic boundaries along the longitude and latitude.

Hazard Quantization is a technique for the optimal selection of a desired number of representative IM maps that can accurately approximate the hazard curve and spatial correlation of the ground motion intensity. In particular, the quality of the approximation is measured imposing mean-square convergence of the selected and weighted maps to the entire IM random field. In other words, HQ does not focus on a few specific statistical parameters (e.g., the probability of exceedance at one site for one value of IM, or the correlation between two sites), but instead aims to holistically capture the probabilistic description of the IM. This main objective of HQ is illustrated in Figure 4.2. Mean-square optimality is particularly valuable for analyses like risk assessment or expected loss prediction. A set of IM maps needs to be available from historical records of past events, or as outcome of the

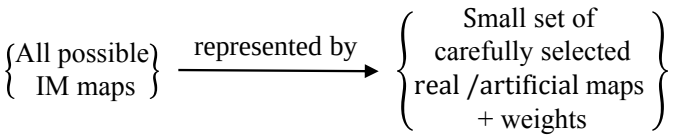


Figure 4.2: The main objective of Hazard Quantization, is to optimally describe in a probabilistic way a ground motion parameter over a region with a small set of selected IM maps and their associated weights.

seismic characterization of the region and proper used of a GMPE (including the residuals). HQ considers these IM maps as realizations of a two-dimensional random field as input and provides an optimal representation with a small number of weighted IM maps as output. This means that for a fixed number of scenarios to be analyzed, HQ provides the set of IM maps and weights that most accurately represents the whole set of potential IM maps.

Hazard Quantization builds upon the theoretical basis of a technique called Functional Quantization by Infinite-Dimensional Centroidal Voronoi Tesselation (FQ-IDCVT). The FQ methodology consists of approximating a generic random function F by means of another function F_N. The function F_N is fully described probabilistically by a collection of N samples (i.e., quanta) and their associated weights, all together called quantizer (Christou et al., 2016b). When the correlation length is average-to-large, FQ (and, in turn, HQ) has been proven to work particularly well.

Following subsections provide a step-by-step description of the HQ algorithm to determine representative ground motion IM maps. The algorithm requires initially a large number N_{events} of IM maps which can be provided either by simulation, from a modified historical earthquake catalog or both. The analyst has to define the quantizer size N_{HQ}, which is the (small-to-moderate) number of IM maps that should be determined by HQ and can later be used in the subsequent steps of a regional loss/risk/resilience analysis (Karamlou et al., 2016).

4.3.1 Input large set of IM maps

The expansion of seismic arrays is providing sets of historical data recorded from large regions. Appropriately modified versions of these sets, if available for either the region of interest or a region with similar characteristics, can be used as input for HQ; however, the most common way of obtaining large sets of samples of IM maps is simulation. This can be done following an approach similar to those presented in Ebel and Kafka (1999) and Crowley and Bommer (2006). The seismic characterization of the region should describe with a certain degree of completeness the temporal and spatial distribution of earthquakes which can occur. This should be coupled with a ground motion prediction model that can simulate the ground shaking generated by each earthquake and its associated variability and correlation (e.g., Boore and Atkison, 2008; Jayaram and Baker, 2009).

4.3.2 Quanta Identification

Once the input set of N_{events} IM maps is available, the HQ module is initialized (Step 0), selecting randomly out of the N_{events} maps a subset of N_{HQ} maps to be used as initial trial quanta. The algorithm then consists of an iterative process. At each iteration the N_{events} IM maps are clustered in N_{HQ} groups and each group is associated with a quantum $Q_j(\text{lon},\text{lat})$. Iterations consist of the following steps.

1. Compute the L^2 distance of each of the N_{events} realizations $\{M_i(\text{lon},\text{lat})\}_{i=1}^{N_{events}}$ from each quantum $\{Q_j(\text{lon,lat})\}_{j=1}^{N_{HQ}}$:

$$d_{ij} = \|M_i(\text{lon,lat}) - Q_j(\text{lon,lat})\|_{L^2} = \sqrt{\int_{\text{lat}_{min}}^{\text{lat}_{max}} \int_{\text{lon}_{min}}^{\text{lon}_{max}} [M_i(\text{lon,lat}) - Q_j(\text{lon,lat})]^2 \, d\text{lon} \, d\text{lat}} \tag{4.1}$$

2. Associate to each quantum $Q_j(\text{lon},\text{lat})$ the IM maps for which $d_{ij} < d_{ih}\ \forall\ h \neq j$. In practice, each IM map is associated to its closest quantum.

3. Update each quantum with the average of the IM maps associated with it

$$Q_j(\text{lon,lat}) = \frac{\sum_{k=1}^{N_j} \hat{M}_j^k(\text{lon,lat})}{N_j} \tag{4.2}$$

where $\hat{M}_j^k(\text{lon,lat})$ represents the sample maps associated with quantum $Q_j(\text{lon,lat})$ and N_j is their number.

4. Iterate steps 1-3 until the following functional, called distortion, converges to a stable value:

$$\Delta = \sum_{j=1}^{N_{HQ}} \sum_{k=1}^{N_j} \frac{1}{N_j} \|\hat{M}_j^k(\text{lon,lat}) - Q_j(\text{lon,lat})\|_{L^2}^2 \tag{4.3}$$

The distortion becomes stable and it is minimized when no rearrangements of the N_{events} IM maps into the N_{HQ} clusters occur during the one iteration.

For a pictorial representation of the proposed algorithm the reader can refer to Figure 4.3.

When clustering by exhaustive comparison is performed (as explained in Step 2), the number of iterations to meet convergence can be very high, up to a few hundredths, and it increases as N_{events} increase. Nevertheless, to describe with a certain degree of completeness the spatial and temporal distribution of earthquakes within a region, a large number of N_{events}, at least some tenths of thousands, is required. The number of iterations could be reduced if smarter clustering methods are applied, as described by Gersho and Gray (1991). For more robust results, the N_{events} realizations could be regenerated at each iteration, which makes the technique less sensitive to the initialization of the quanta, but requires additional computational time for the sample IM maps regeneration. In this case, the iterative scheme is terminated when either the reduction of Δ during the past five iterations is smaller than a certain percent or when a maximum number of iterations is completed (Christou et al.,

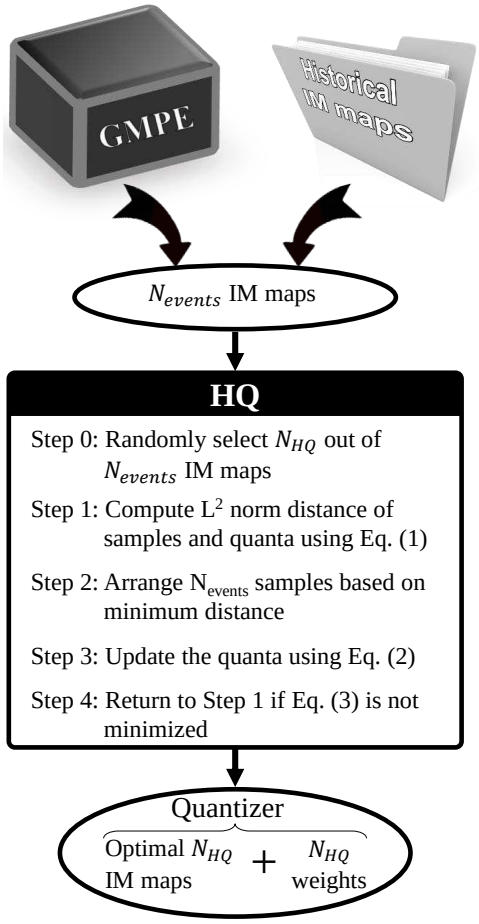


Figure 4.3: Flowchart of the basic steps of the HQ algorithm.

2016b). In the case the parameters in the ground motion prediction model (e.g., earthquake magnitude) for the simulation of the N_{events} IM maps are generated through variance reduction techniques (e.g., importance sampling), then the resulting corresponding sample weights should be taken into account throughout the procedure, in particular Equations (4.2)-(4.4).

4.3.2.1 Output

When the HQ algorithm has converged, the methodology provides a set of representative IM maps (i.e., the final set of quanta $\{Q_j(\text{log,lat})\}_{j=1}^{N_{HQ}}$) along with their associated weights $\{p_j\}_{j=1}^{N_{HQ}}$. In particular, the probability associated with each IM map $Q_j(\text{log,lat})$ is computed as:

$$p_j = \frac{N_j}{N_{events}} \tag{4.4}$$

The hazard-consistent annual probability of occurrence of an IM map can be computed by using a Poissonian model, and thus dividing Equation (4.4) by the annual occurrence rate of any event:

$$p_{annual_j} = \frac{p_j}{\lambda_{annual}} \tag{4.5}$$

If, for a specific application, it is preferable to use a set of real maps rather than the quanta, then the analyst can replace each quantum with its closest map according to the L^2 distance.

4.4 Alternative techniques for the selection of representative IM maps

A direct comparison of results obtained by HQ and two techniques will be provided as part of the numerical applications. First, the "optimization-based probabilistic scenario" (OPS) method will be described (Han and Davidson, 2012; Manzour et al., 2016). The second

methodology is part of the approach presented by Jayaram and Baker (2010). This approach includes several modules, such as importance sampling of earthquake magnitudes and residuals. These modules can be applied also to HQ and OPS, and are particularly useful for the assessment of specific probabilistic metrics, such as the reliability of distributed systems, but not as much for general studies like expected loss estimation or risk analysis. The actual selection of IM maps is independent of these modules, and Jayaram and Baker (2010) proposed to do it using a methodology called "*k*-means" clustering.

4.4.1 Optimization-based probabilistic scenario method

This method uses mixed integer linear programming optimization to select first a reduced set of median IM maps (OPS-1), and then a reduced set of IM maps including residuals (OPS-2). The optimization formulation is the same for both selection steps. The objective of the optimization is to make the regional hazard estimated by the reduced set of maps match as closely as possible the reference hazard at a selected set of locations and return periods. It should be emphasized that this objective does not account for the spatial correlation of the IM values. The overall correlation is induced solely by the intrinsic correlation embedded in each map.

4.4.2 *k*-means clustering

k-means clustering (MacQueen, 1967) is used by Jayaram and Baker (2010) as a data reduction technique to develop a small catalog of ground motion IM maps. The *k*-means algorithm is the same as Lloyd's algorithm (Lloyd, 1982), which served as a starting point for the development of FQ-IDCVT and, in turn, HQ (Miranda and Bocchini, 2015a). The main difference is that HQ uses the quanta to represent the clusters, while Jayaram and Baker pick a random map from each cluster.

k-means works with vectors: $\mathbf{M_k}$ is a vector of values taken by the IM map at a finite set of locations (e.g., important facilities, or a regular grid) and $\mathbf{Q_j}$ is the centroidal vector of cluster j, which can be interpreted as a discrete version of a quantum $Q_j(\text{lon,lat})$. The discrepancies among IM maps in the *k*-means technique are usually measured by the Euclidean distance and the clustering is performed through the minimization of the following metric:

$$U = \sum_{j=1}^{N_{kmeans}} \sum_{\mathbf{M_k} \in \mathbf{G}^j} \|\mathbf{M}_k - \mathbf{Q}_j\|^2 \tag{4.6}$$

where $N_{kmeans} = N_{HQ}$ is the number of clusters and $\mathbf{G^j}$ denotes the set of maps that belong in cluster j.

The steps followed by the *k*-means algorithm are the same as those presented for HQ, but with discretized versions of the maps. Therefore, there are two important similarities between *k*-means clustering and the HQ technique: First, both techniques cluster maps in groups whose members have similar characteristics and, secondly, both algorithms use an iterative refinement to group maps in clusters using comparable discrepancy metrics.

On the other hand, HQ and the methodology proposed by Jayaram and Baker (2010) have a different genesis that results in important differences. First, HQ stemmed from methodologies developed for random functions. For this reason it naturally treats continuous functions over a space, and therefore accounts for the value of the IM at all infinite sites of the region. *k*-means was developed for vector clustering, and thus considers IM values at a discrete set of locations. Second, HQ uses the quantum as representative map from each cluster (or the map closest to the quantum), which is proven to be an optimal choice, whereas Jayaram and Baker (2010) suggest to pick a random IM map from each cluster.

4.5 Application: Regional seismic hazard of the Charleston seismic zone

In this section, HQ is applied to the Charleston seismic zone to demonstrate the utility of proposed framework for regional probabilistic hazard analysis. Charleston and the surrounding area are prone to multiple hazards, including earthquakes, and have dense infrastructure networks, making it a good case study for regional hazard analysis. The regional hazard is calculated with HQ and compared to three other techniques: conventional MCS to have a reference solution; k-means clustering (Jayaram and Baker, 2010); and OPS (Han and Davidson, 2012).

4.5.1 Charleston seismic characterization

The input data and the probabilistic description of the parameters for the seismic characterization of the Charleston region are based on the Central and Eastern United States Seismic Source Characterization Project (CEUS-SSCn, 2012). The seismic sources of the Charleston area that are used in the 2014 USGS National Seismic Hazard Mapping Project (Peterson et al., 2014) are adapted from the CEUS-SSCn.

The CEUS-SSCn project follows an event-tree procedure for the complete seismic characterization of Charleston's repeated large-magnitude earthquake (RLME) source. This procedure includes spatial and temporal information and descriptions of fault geometry and style, among other parameters. In this application, the CEUS-SSCn report is used for spatial characterization, while the USGS information is considered for the recurrence uncertainty.

Geological surveys show that the available historical data for the spatial seismic char-

acterization of the Charleston region do not identify any causative faults. For example, neither the prehistoric earthquakes nor the 1886 earthquake is considered to belong to any specific fault or fault zone. However, the Charleston region is associated with a pattern of observed seismicity, which, based on the field evidence, can be characterized as a RLME source. In particular, for the Charleston region, RLME sources are locations of repeated earthquakes with magnitude M$\geq$6.5 in the historical and paleoearthquake records.

The predicted Charleston earthquakes are expected to occur at uniformly distributed locations within local, narrow, and regional area sources (or zones), represented in Figure 4.4(a). The local Charleston source is given a weight of 0.50 in the model and the predicted ruptures have a NE orientation, parallel to the zone's long axis. The local source has strict boundaries, such that the ruptures are not allowed to extend beyond them. Similarly, the narrow Charleston source, receiving a weight of 0.30, is modeled as an area of distributed faults or active tectonic features instead of a particular fault. In the narrow model, the orientation of the predicted ruptures is parallel to the zone's long axis and in this case their lengths are allowed to extend beyond the NE and NW boundaries. The regional zone, which also has strict boundaries, is the largest zone in the Charleston area, approximately 260 x 150 km, and envelops both the local and narrow zones. In the current model, the regional source configuration is given a 0.20 weight, and its predicted orientations are either parallel to the long axis of the zone with probability 0.80 or oriented perpendicular to the long axis of the zone with probability 0.20.

Figures 4.4(b)-(d) depict the distribution of 100 sample event faults in the three Charleston RLME source zones. In this application, the predicted ruptures in all the Charleston sources are modeled as occurring on vertical strike-slip faults. The shear-wave velocity, or soil type, in the upper 30 m of the crust is $V_{S30} = 760$ m/s, given that the wave travels through a firm rock. It should be noted that considering a uniform soil condition in the study region reduces the variability of the IM maps and introduces an approximation; however, this is a

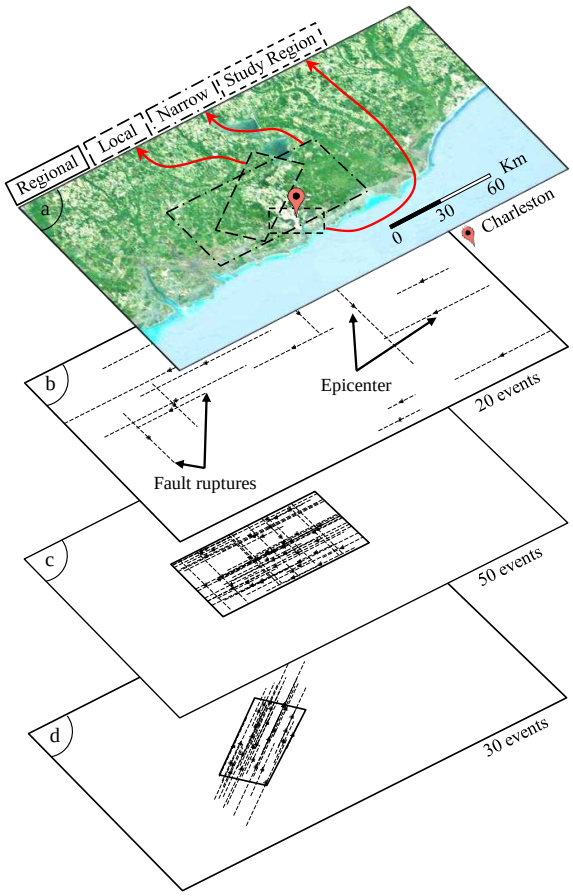


Figure 4.4: Distribution of 100 scenarios (i.e., epicenters with their corresponding fault rupture lengths) generated by plain MCS from each of the Charleston RLME source zones. 16 and 4 simulated epicenters with their corresponding fault ruptures parallel and perpendicular to the Regional zone's long axis, respectively (b), 40 and 10 simulated epicenters with their corresponding fault ruptures parallel and perpendicular to the Local zone's long axis, respectively (c) and 30 simulated epicenters with their corresponding fault ruptures parallel to the Narrow zone's long axis (d).

simplification that does not interfere with the objectives of this research.

The magnitudes of the predicted large earthquakes in the Charleston RLME zone are based on the assessment of the various sources of available data and modeled with a uniform distribution with bounds and weights indicated in Table 4.1.

Sequences of earthquakes within the Charleston seismic zone are usually spaced about 500 to 1000 years apart, but sometimes quiescent intervals can last a few thousand years

Table 4.1: Expected Charleston RLME magnitudes, and lower and upper bounds of the associated uniform distributions based on CEUS-SSCn (2012).

Type	Expected Charleston RLME (M) (M)	Weight	Lower bound (M)	Upper bound (M)
1	6.7	0.10	6.46	6.95
2	6.9	0.25	6.65	7.15
3	7.1	0.30	6.85	7.35
4	7.3	0.25	7.05	7.55
5	7.5	0.10	7.25	7.75

(CEUS-SSCn, 2012), making predicted earthquake recurrence highly uncertain and highly variable through time. Following the USGS provisions, this application considers a return period of 529 years and uses the time-independent Poisson model for the recurrence of an event.

For the generation of the surface rupture length of an earthquake event, the following equation is used (Department of Homeland Security, 2003; Wells and Coppersmith, 1994):

$$log_{10}(L) = a + b \cdot M \tag{4.7}$$

where L is the rupture length (km), M is the moment magnitude of the earthquake and a, b are regression coefficients defined as $a = -3.55$ and $b = 0.74$. Finally, the hypocenter depth, (i.e., the point at which the rupture begins and the first seismic waves originate) is modeled with a uniform distribution between 5 and 20 km.

Ground motion prediction model

The ground motion intensities are simulated based on the GMPE:

$$ln(Sa_{(T=1.0s)}(\text{lon,lat})_i) = ln(\overline{Sa_{(T=1.0s)}(\text{lon,lat})}_i) + \varepsilon(\text{lon,lat})_i \cdot \sigma_i + \eta_i \cdot \tau_i \tag{4.8}$$

where $Sa_{(T=1.0s)}(\text{lon, lat})_i$ denotes the i-th IM map of the two-dimensional spectral acceleration (Sa) at period $T = 1.0$ sec, "lon" and "lat" denote the longitude and latitude coordinates in the spatial domain, $\overline{Sa(\text{lon, lat})}_i$ denotes the i-th realization of the predicted median value of the ground acceleration obtained from the empirical ground motion prediction model, and the products $\varepsilon(\text{lon, lat})_i \cdot \sigma_i$ and $\eta_i \cdot \tau_i$ are the i-th realizations of the intra- and inter-event residuals, respectively. Specifically, parameter $\varepsilon(\text{lon, lat})_i$ denotes the i-th realization of the normalized intra-event residual, modeled as a normal random field with zero mean and unit standard deviation. The scalar parameter σ_i denotes the i-th realization of the standard deviation term and scales $\varepsilon(\text{lon, lat})_i$. It is estimated as part of the empirical GMPE model and is a function of the period T of interest. Parameter η_i is constant over the region, and is the normalized inter-event residual, modeled as a normal random variable with zero mean and unit standard deviation. Finally, the scalar parameter τ_i is standard deviation of the inter-event residual and is likewise calculated as part of the empirical GMPE.

This application uses Boore and Atkinson's 2008 empirical GMPE to calculate the median $ln(\overline{Sa(\text{lon,lat})}$ term, σ and τ. The normalized intra-event residual ε is computed based on the work presented by Jayaram and Baker (2009). Their predictive model for spatial correlation constructs the covariance of ε. The correlation between the residuals at two sites is a function of the distance h between the sites, the period of interest and the soil condition similarity:

$$\rho_{\varepsilon(\text{log,lat}),\varepsilon(\text{log,lat})}(h) = exp(-3h/b_\varepsilon) \tag{4.9}$$

where b_ε denotes the rate of decay of spatial correlation, and thus the ratio $3/b_\varepsilon$ is proportional to the correlation length. In this application, parameter b_ε equals 25.7 for $T = 1.0s$.

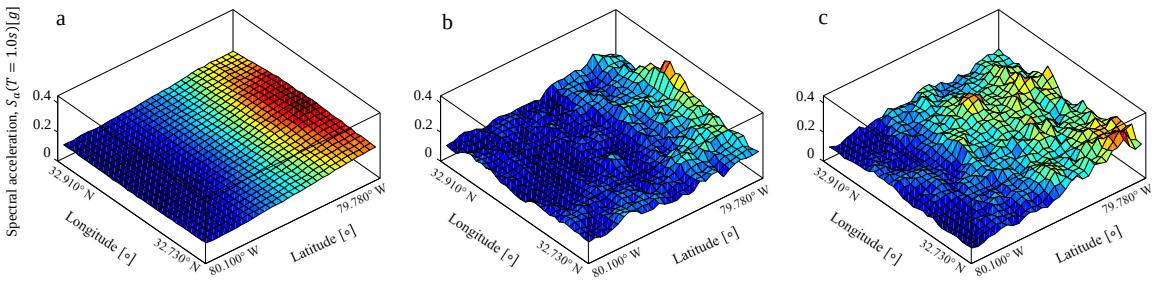


Figure 4.5: IM maps resulting after performing HQ on the study region with quantizer size equal to 135. One of the 135 quantum IM maps (a), the closest IM map in L^2 distance to the quantum (b) and a random IM maps selected from the cluster of IM maps which the quantum represents (c).

4.5.2 Parameters of the algorithms

Each hazard representation methodology requires the selection of input values for its parameters, and in this application, considerable effort was put into selecting parameters for each technique in a way that allowed for a fair comparison among them. Some parameters are specific to a certain methodology, such as OPS's return periods of interest, whereas others are common to all of them, such as the target number N of desired IM maps. Even common parameters may be used in different ways by the various techniques. N, for instance, is taken as the exact number of IM maps to be selected by HQ and k-means, whereas it serves as maximum number of IM maps for OPS.

In the OPS method, 3,000 sample median IM maps ($N_{eventsOPS1}$) were generated as input for OPS-1, which then selected no more than 500 (N_{OPS1}) output median maps and their corresponding annual occurrence probability. Before OPS-2, the samples of the intra- and inter-event residuals were generated and then superimposed to the median maps from OPS-1. For this application, the sample intra- and inter- event residual maps were paired to each of the median maps based on their annual occurrence probabilities. As a result of this pairing, OPS-2 receives as input $N_{eventsOPS2} = 3{,}000$ IM maps and the target number of outcome maps (N_{OPS2}) is set to 300. A 6 x 6 grid encompassing the geographic coordinates

$32,730^{\circ}$ N - $32,910^{\circ}$ N (latitude range) and 79.780 W - 80.100 W (longitude range) was selected as the resolution in both OPS-1 and OPS-2. The selected return periods for the two phases of OPS are 800; 1,200; 2,500 and 5,000 years, and the mixed-integer programming optimizations were performed using the Gurobi solver (Manzour et al., 2016). While it is not possible to relate directly N_{events} with the pair of parameters $N_{eventsOPS1}$ and $N_{eventsOPS2}$, in some sense it could be said that with these parameters OPS investigated a set of 3,000 x 3,000 = 9 million IM maps.

For the HQ method the sample space of all possible realizations is defined by N_{events} IM maps and for this application it was set equal to 50,000. The IM maps are discretized on a 30 x 30 point grid, resulting in a study region mesh of 1 km x 0.7 km. The number N_{HQ} of reduced IM maps was set to 135, because this was the average number of IM maps eventually selected by OPS. In this application, there is no sample regeneration during the iterations of the HQ algorithm and clustering is performed by exhaustive comparison without the use of smart clustering techniques.

Similarly, in the *k*-means methodology number $N_{eventsKmeans}$ of IM vectors was 50,000 (like N_{HQ}) and the IM values were computed at the same 30 x 30 grid. Also, the number of outcome IM maps equals the choice made for HQ: $N_{Kmeans} = N_{HQ} = 135$. Therefore, for a fair comparison, the discretized $\{M_i(\text{lon,lat})\}_{i=1}^{N_{events}}$ and $\{\boldsymbol{M}_i\}_{i=1}^{N_{eventsKmeans}}$ are the same. Furthermore, the initial sets $\{Q_j(\text{lon,lat})\}_{j=1}^{N_{HQ}}$ and $\{\boldsymbol{Q}_j\}_{i=1}^{N_{kmeans}}$ for HQ and *k*-means to initiate the iterative procedures were always selected to be the same. This guarantees that HQ and *k*-means iterative algorithms perform the same number of iterations and yield the same clusters.

The generation of the initial IM maps was performed through entirely random sampling of their parameters (e.g., earthquake location, magnitude, fault orientation, residuals). Importance sampling or other sampling techniques could have been used, but in this case there

was no reason to focus on any particular type of seismic event.

An important characteristic of all methods is the number of locations considered in the IM maps. In the OPS and *k*-means methods a set of specific geographic locations should be identified, sometimes on a regular grid. For example, they can include the locations of significant assets in an infrastructure network. For HQ, it is more natural that the discretization of the IM random field is made on a regular grid with a resolution that captures the spatial correlation for the points of interest. In the case of seismic hazard analysis, a discretization of the IM grid with approximately 1 km resolution is considered adequate for applications in civil engineering (Jayaram and Baker, 2010; Han and Davidson, 2012; Manzour et al., 2016).

Each method was run 100 times with different inputs. For HQ and *k*-means 100 sets of 50,000 IM maps were generated, and for OPS the procedure was run 100 times with different sample events for the OPS1 and OPS2 phases. The solutions obtained from HQ, OPS and *k*-means are compared with the reference solution obtained by MCS with 500,000 samples and the results are evaluated to showcase the benefits and drawbacks of each method.

4.6 Results and discussion

The product of HQ, when one analysis was run, was a set of 135 IM maps and their associated weights. Figure 4.5(a) shows the representative map from one of the 135 partitions of the sample space. Similarly, the outcome of *k*-means after the first analysis was a set of 135 IM maps and their associated weights. An example of one of the IM maps resulting from the *k*-means technique is depicted in Figure 4.5(c). The result of the first of 100 analyses, when the OPS was used, was a set of 144 IM maps and their corresponding annual probabilities of occurrence. The outcome of any of these methodologies can later be used as input in following regional loss estimation analysis steps. The IM at certain locations

extracted from one of the resulting IM maps can be used as input of a structural simulation to calculate demands parameters which characterize the structure's response in terms of acceleration, for instance. This structural response will then be weighted based on the weight of the input IM map.

In the OPS method, after 100 analyses have been run, an average of 8 median IM maps were selected at the end of OPS-1 and an average of 135 IM maps were selected at the end of OPS-2. Neither the output of OPS-1 nor of OPS-2 could be controlled precisely by parameters N_{OPS1} and N_{OPS2}. As such, the analyst does not benefit from the selection of a number of median IM maps or IM maps with residuals that are greater than 8 or 135 and smaller than the $N_{eventsOPS1}$ and $N_{eventsOPS2}$, respectively. For example, in this application the average number of IM maps selected by OPS does not change increasing N_{OPS2} until it reaches the value 800. This means that the methodology cannot find a combination of IM maps of number between 135 and 800 that is better than what can be achieved with only 135 IM maps.

The first quantitative comparison is done on the ability to correctly represent the hazard at all locations of the region, considered individually. The comparison is done at the locations of the 30 x 30 grid in the study region, and focuses on the values of the hazard curve at eight return periods (i.e., T = 800; 1,000; 1,200; 1,600; 2,000; 2,500; 4,000 and 5,000). This choice has been made because according to FEMA 366 (FEMA, 2008), loss estimates can be generated with as few as eight return periods. For instance, Figure 4.6 shows the reference and the approximated hazard curves at location 32.760°N, 79.830°W as obtained from one run (i.e., one random seed number). This geographic location belongs also to the 6 x 6 grid considered by the OPS method. The eight circles represent the hazard related to eight selected return periods. When 135 IM maps are considered, this picture shows that all three methods provide a good estimation of the hazard at the selected return periods. Similar results on the estimation of the hazard curve are obtained for other locations within

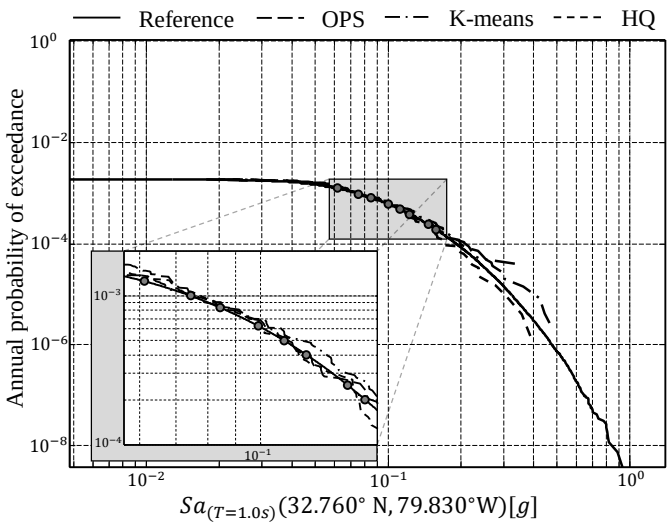


Figure 4.6: Hazard curve plot in logarithmic scale obtained with reference MCS, OPS and quanta samples at location 32.760°N, 79.83°W.

the region.

The error in the hazard curve in terms of the spectral acceleration at $T = 1.0s$ ($Sa_{(T=1.0s)}$) is defined as (Han and Davidson, 2012):

$$\frac{Q^r(lon_l, lat_k) - \tilde{Q}^r(lon_l, lat_k)}{\tilde{Q}^r(lon_l, lat_k)} \tag{4.10}$$

where r is the return period index, $Q^r(lon_l, lat_k)$ and $\tilde{Q}^r(lon_l, lat_k)$ are the assessed and reference ground motion intensities for a site with coordinates (lon_l, lat_k) and return period r, respectively. The overall hazard curve error over the study region is captured by the mean over all sites (lon_l, lat_k) and all return periods r_m:

$$E_{HC} = \frac{1}{n_r n_{lon} n_{lat}} \sum_{r=1}^{n_r} \sum_{l=1}^{n_{lon}} \sum_{k=1}^{n_{lat}} \left| \frac{Q^r(lon_l, lat_k) - \tilde{Q}^r(lon_l, lat_k)}{\tilde{Q}^r(lon_l, lat_k)} \right| \tag{4.11}$$

where n_r, n_{lon} and n_{lat} are the upper bounds of the indexes running over the return periods, the longitudinal axes and the latitudinal axes respectively, in this case $n_r = 4$, $n_{lon} = 30$ and $n_{lat} = 30$.

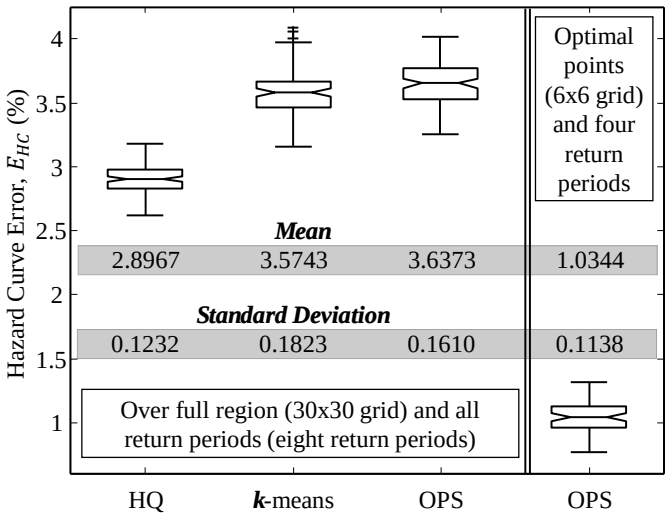


Figure 4.7: Box and Whisker plots of the E_{HC} when 100 experiments are conducted for each method: OPS, HQ and k-means. The E_{HC} is evaluated at the 30x30 grid points and at the eight return periods. Additionally, the E_{HC} obtained by OPS when it is evaluated at the 6x6 selected points of the grid and for four return periods is shown for the same 100 experiments at the last column.

Figure 4.7 shows box and whisker plots of the E_{HC} as defined in Equation (4.11) for each method. As mentioned, the analyses have been repeated 100 times using different random seeds and thus Figure 4.7 provides an overview of the stochastic error on the hazard curve for each method. The last box depicts the E_{HC} evaluated for the reduced set of IM maps yielded by OPS only at the points and return periods considered for the optimization. The objective function in OPS is to minimize precisely E_{HC} at these points and return periods and thus, when only these specific points and return periods are considered, OPS provides better results compared to the other methods. OPS, however, is constrained by the problem's dimensionality because the computational time and resources needed to optimize the objective function increase non-linearly as resolution increases. It would not have been possible to expand the OPS method to optimize a 7 x 7 grid with the available resources. The E_{HC} errors obtained by HQ evaluated at the 30 x 30 grid points and at eight return periods are slightly smaller than those yielded by the other methods, showing that HQ accurately approximates the hazard at each site as well as these established methodologies.

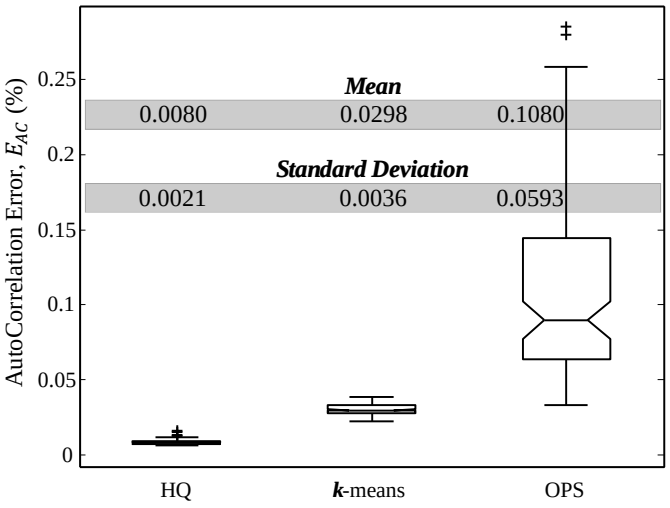


Figure 4.8: Box and Whisker plots of the E_{AC} when 100 experiments are conducted for each method: OPS, HQ and k-means. The E_{AC} is evaluated at all combinations of pairs of geographic locations.

After assessing HQ's ability to capture the hazard at each individual site, it is important to assess the ability of HQ to correctly capture the correlation among different sites. To accomplish this, an estimation of the autocorrelation was computed between every pair of locations in the grid used to represent the IM map. If N_{HQ} values of IM for each one of the discrete locations of the grid at coordinates (lon_l, lat_k) are available, then the autocorrelation is computed as:

$$R_{QQ}(lon_l, lat_k, lon_h, lat_g) = \sum_{j=1}^{N_{HQ}} Q_j(lon_l, lat_k)\, Q_j(lon_h, lat_g)\, p_j \tag{4.12}$$

where

$$\begin{aligned} lon_l &= \text{lon}_{min} + (l-1)\cdot\Delta\text{lon} & l &= 1,\ldots,n_{lon} \\ lat_k &= \text{lat}_{min} + (k-1)\cdot\Delta\text{lat} & k &= 1,\ldots,n_{lat} \\ & & lon_{n_{lon}} &= \text{lon}_{max} \\ & & lat_{n_{lat}} &= \text{lat}_{mat} \end{aligned} \tag{4.13}$$

and p_j is defined in Equation (4.4). As Equation (4.12) shows, the autocorrelation is a 4

dimensional surface, so it cannot be visualized. The overall error for all pairs of geographic locations can be expressed as the mean error over all pairs of sites:

$$E_{AC} = \frac{2}{(n_{lon} \cdot n_{lat})^4 + (n_{lon} \cdot n_{lat})} \cdot \sum_{l=1}^{n_{lon}} \sum_{k=1}^{n_{lat}} \sum_{h=1}^{l} \sum_{g=1}^{k} \left| \frac{R_{QQ}(lon_l, lat_k, lon_h, lat_g) - \hat{R}_{QQ}(lon_l, lat_k, lon_h, lat_g)}{\hat{R}_{QQ}(lon_l, lat_k, lon_h, lat_g)} \right| \tag{4.14}$$

Figure 4.8 shows box and whisker plots of the autocorrelation error for all considered methods. In this particular example, E_{AC} is small for all methods. The error associated with HQ is negligible and, furthermore, it is up to one order of magnitude smaller than the error obtained from OPS. The figure also highlights HQ's near-zero error estimator variance.

Based on the way HQ and k-means have been implemented in this application, they differentiate only in the last step of the algorithm, in which IM maps are selected to represent each cluster. It is important to mention that a higher number of N_{events} sample IM maps (e.g., 500,000) would improve the HQ results. Mean square convergence of the ensemble of representative samples to the desired IM field is satisfied for a high number of N_{events} realizations. The higher this number, the better the results on the E_{HC} and E_{AC} errors for HQ. Conversely, for k-means, a higher N_{events} may have adverse effects, as randomly selected IM maps from each cluster with more samples are less likely to be the optimal representative sample.

Another important consideration is the possible limitation imposed by the computational time and resources which OPS requires. All the analyses were run on a 12-core CPU with 64MB of RAM and each run of the OPS-2 finished in less than one hour when a 6 x 6 grid and four return periods were considered. Any increase in the resolution of the IM maps or in the number of sample IM maps in OPS-2 would result in a much longer computational time and it would reach the limit of these specific computational resources.

Table 4.2: Comparison and contrast table of the features of the frameworks discussed. The gray hatch cells indicates the superiority of one framework over the others for a specific characteristic.

	Framework		
Features	HQ	K-means	OPS
Selects exactly the desired number of IM maps	✓	✓	Not always
Input IM maps size limits methodology			✓
Ease of implementation	Easiest	Complex if IS is implemented	Complex
Requires external optimization solver			✓
Accurate hazard curves estimate for all locations	Accurate	Accurate	Accurate[1]
Can select IM maps targeting specific hazard levels			✓
Requires multiple selection steps			✓
Accurate on regional autocorrelation estimate	Most accurate	Accurate	Less accurate
Was developed to work together with specialized simulation techniques		✓	✓
Robustness (σ in Fig. 4.8)	0.0021	0.0036	0.0593

[1]Most accurate when the locations of interest are those used for the optimization.

For example, an increase in the resolution of the IM maps to a grid of 7 x 7 points took more than 15 hours for the optimization to reach convergence and thus the analyses were terminated. Therefore, the coarser grid of 6 x 6 points for each of the 3,000 IM maps and the four return periods were selected in OPS-2.

Table 4.2 compares and contrasts the features of the three discussed frameworks. Some aspects have already been discussed, therefore only the features which have not yet been reviewed are analyzed next.

Ease of implementation: The framework presented by Jayaram and Baker, which considers *k*-means for the selection of the IM maps, uses importance sampling for the generation of a few of the GMPE's parameters. Similarly, OPS requires advanced sampling techniques (stratified sampling) for the same parameters and some preliminary knowledge on the use of optimization techniques, making this technique more challenging to implement

in practice. Additionally, OPS necessitates a two-stage IM maps generation and selection process. HQ, on the other hand, requires only a purely random sampling of the parameters for the generation of the IM maps. Like *k*-means, in the selection phase, only basic operations are needed to code the iterative algorithm, making this technique the easiest of the three to implement.

Requires solver: The solution of the two optimization formulations in the OPS framework requires the use of an appropriate mixed-integer linear programming solver. Depending on which solver is chosen, the results may be different. In contrast, HQ and *k*-means do not need a solver, which makes these frameworks less dependent on external software whose calculations may not be fully controlled by the analyst.

Accurate regional hazard estimate: Based on the results presented in Figure 4.7, HQ provides the best regional hazard estimate for this application. OPS provides accurate estimates in this application even though the regional hazard was evaluated at denser grid points than the 6 x 6 grid which it used in OPS-1 and OPS-2. If the regional hazard estimate is evaluated only at the locations used in OPS, then the average error for all the points' hazard curves (at the selected return periods) is minimized. In this case, OPS yields the most accurate results.

Selects IM maps targeting specific hazard levels: One of the input parameters in the OPS optimization formulation is the identification of the specific return periods at which the hazard curve error is minimized, which are equivalent to the probability of occurrence of interest. These return periods can be arbitrarily defined, therefore specific hazard rates can be selected to be analyzed, such as only rare events with high return periods. In other words, the error minimization can focus on the upper tail of the cumulative distribution. Conversely, HQ does not minimize the hazard error at specific points on the hazard curve. It focuses on the main bulk of the distribution, making it ideal for regional loss analyses.

Robustness: Figure 4.8 shows that standard deviation of the autocorrelation error E_{AC} obtained from HQ is at least 50% smaller than the corresponding σ yielded from k-means. The standard deviation on the E_{AC} resulting from OPS is one order of magnitude higher than HQ and k-means. In conclusion, this makes HQ the most robust framework among k-means and OPS for the selection of a reduced set of IM maps.

4.7 Concluding Remarks

A novel methodology called Hazard Quantization is presented for the selection of a set of optimal stochastic IM maps representing the regional hazard over a geographic region. Hazard Quantization treats the regional IM maps as realizations of a two-dimensional non-Gaussian and non-homogeneous random field. The method imposes mean-square convergence of the ensemble of representative samples to the desired IM field. HQ is applied to the regional seismic hazard analysis of the Charleston, South Carolina region. Other popular techniques are reviewed and applied in this example for comparison.

The results of this application showcase this technique's versatility and accuracy, and, most importantly, they showcase that it is an easily implementable, robust technique that can be used without requiring specialized simulation techniques or hierarchical parameter sampling. These features can make HQ particularly appropriate and appealing to accelerate the process that is bringing regional hazard analysis to engineering practice.

Chapter 5

Hurricane simulation framework of synthetic directional wind speeds for hazard and risk analysis in the United States

5.1 Introduction remarks

Hurricanes are among the most destructive natural disasters. A hurricane is a tropical cyclone that occurs in the Atlantic ocean and in this chapter the term hurricane is used interchangeably with tropical cyclone and storm. A Category 4 or 5 hurricane landfalling in the South or East Coast of the United States lowers the U.S production (EIA, 2018), increases unemployment (Brown, 2006), and, even smaller storms depress the stock market (Finance.yahoo, 2017). The United States is vulnerable to hurricane damage due to its population and infrastructure distribution. Coastal shoreline counties create 40% of America's jobs and they are responsible for the 45.6% of its gross domestic income (NOAA, 2017). Moreover, despite the decrease of the country's death toll from hurricanes over the years (Goklany, 2009; Blake et al., 2011) due to improved forecasting systems, improved structures with higher load resistance, and the ability to provide timely warning to public, there has been a significant increase in economic losses. Four of the six costliest U.S. tropical cyclones occurred in the past 7 years as depicted in Table 5.1 (NHC/NOAA, 2018).

Wind damage from hurricane contributes significantly to the cost among other sources (e.g., storm surge, precipitation). The Congressional Budget Office (CBO) estimates an expected annual economic loss of $55 billion due to hurricane wind and storm-related

Table 5.1: Costliest landfalling U.S. tropical cyclones between 1900-2018 in billion dollars of damage (not adjusted for inflation) (NHC/NOAA, 2018).

Rank	Name	Year	Cost(B$)	States	Death toll (US only)
1	Katrina	2005	125	SE, FL, LA, MS	1,245-1,836
2	Harvey	2017	125	TX, LA	106
3	Maria	2017	90	SE, FL, LA, MS	7
4	Sandy	2012	65	NY, NJ, PA, MA	157
5	Irma	2017	50	FL	96
6	IKE	2008	30	TX, LA	113

flooding on the residential, commercial and public sectors. The wind source damage alone comprises 41.2% of the total annual expected cost (CBO, 2019).

To reduce the hurricane wind induced risk, engineers work to build resilient structures and infrastructure systems (Bocchini et al., 2019) to reduce short and long term losses, and to accelerate the community's socio-economic recovery after a natural disaster (reduce long term losses). However, thus far emphasis has been given in modeling and prediction of the hurricane wind intensity. Recently published studies demonstrate that accurate assessment of structural behavior—a predictor of structural damage and risk—under wind load requires reliable information about both wind intensity and directionality (Mara and Hong, 2013; Gao and Wang, 2018; Ma et al., 2019).

Until the late 1990s, the mathematical simulation of hurricanes were exclusively based on single site probabilistic models and sometimes it is still being used during the past recent years. Examples of such models used for risk analysis applications in the U.S. including: Neumann (1991); Georgiou et al. (1983); Georgiou (1986); DeMaria et al. (2006). These models were based on the pioneer work of the hurricane simulation approach introduced by Russell (1969). They require site-specific probability distributions to model key parameters of hurricane mathematical models, such as the radius to maximum wind, central pressure deficit of the approaching hurricane, hurricane heading angle, translation speed,

and the closest crossing distance of the hurricane's eye. These variables are created using historical data from the specific site. After the probability distribution has been constructed, samples are drawn using Monte Carlo approaches, and the hurricane's path is modeled as a straight line. The hurricane's intensity is assumed to be constant until the storm makes landfall and then it starts decaying based on a filing rate model. During this process, wind field models are utilized to estimate the wind speed intensity. The most notable differences of the various site-specific models are concentrated on the physical and statistical models used. For example, different physical models can be found in the various wind field models, and different statistical models can be found in the filling rate models, the probability distributions used to fit the historical data, and the models used to approximate the distance of the storm's center form the site.

A probabilistic approach for the hurricane hazard assessment was introduced by Vickery et al. (2000a), who first proposed to generate synthetic tropical cyclone tracks. In his pioneer work, the full track of a tropical cyclone and its intensity are modeled, starting from the storm genesis over water, its propagation over water and land (for landfalling storm), and ending with its dissipation. In contrast to previous models that were limited to a specific site or a small, confined region, Vickery's stochastic model was innovative in that it modeled hurricane risk along the North American continent's entire coastline. Various models based on Vickery's stochastic model have been proposed to accurately predict hurricane occurences, storm size, landfall locations, and maximum windspeeds; however, very little progress has been made in estimating expected hurricane damage. This is for three reasons: First, very few academic research groups conducting research in the field or insurance and reinsurance companies have put together such simulators. Thus, accessing proprietary data is untenable. Second, the published documentation does not always provide the required information for implementing and putting together a stochastic hurricane simulator. Third, and mostly importantly, the current models predict risk based on the

maximum wind intensity, but fail to take into consideration joint wind and directionality.

This chapter proposes a framework for simulating synthetic directional hurricane wind speeds. The outcome of the simulator can be used to construct joint probability density functions (PDFs) of wind intensity and directionality along the Gulf of Mexico, East Coast of the U.S. and offshore. The joint PDFs and the marginal wind directionality distribution, displayed in polar PDFs, can be utilized for site-specific hazard analysis and long-term risk studies. In addition, the data can be used for event based short-term studies such as post-event damage assessments.

For the simulation of joint hurricane wind intensity and wind directionality, the hurricane track modeling approach is adopted. The modeling of wind speed intensities and directionalities requires the implementation of many modules, which will be described below. Each of these modules is used to model a characteristic aspect of the hurricane simulator. Together, the modules combined, produced accurate directional wind speeds. In the rest of the chapter, each modules is presented and implementation details are discussed.

5.2 Hurricane track modeling

The currently documented hurricane track simulation models are reviewed in summary next and in a chronological order.

The first full-scale hurricane track model was introduced by Vickery et al. (2000a). The authors used a multiple linear regression analysis to model the storm's translation speed and heading angle in each of the 5° x 5° grids covering the entire Atlantic Basin, and the regression coefficients were determined separately for westerly and easterly headed storms. During the storm's track simulation process the starting values of the variables describing the storm's track are sampled from the existing historical HURDAT database

(Jarvinen et al., 1984). Knowing the storm's position, time, speed and heading angle at time t_i, the translation speed at time t_{i+1} can be determined. Similarly the storm's heading angle at time t_{i+1} is determined as a function of the storm's location, translation speed and heading angle at time t_i and heading angle at time t_{i-1}. This model has been used in many of the proprietary models developed by insurance companies to predict losses due to hurricanes (AIG personal communication, February 2019). Despite its widespread use, the track's propagation does not consider properly the local topology. Rather, it depends exclusively on historical data that are discretized in $5^\circ \times 5^\circ$ geographic grids. That is to say, a hurricane track that barely crosses the boundary of an adjacent geographic cell does not take into account half of its local topological information, but only relies on the information included in the current geographic cell.

Other researchers have since adopted this modeling approach. Casson and Coles (2000) proposed a simplistic approach where the simulated tracks are sampled from the historical catalog and a random perturbation is superimposed uniformly throughout the simulated track's path. The authors proposed a perturbation distributed as $N(0, \sigma^2)$, where σ=100 nautical miles. The authors, however, did not provide comparisons between the spatial directional distributions and translation speed distributions of both the historical and simulated hurricane tracks.

In 2005, a modeling approach based on the Markov Chain was introduced (Powell et al., 2005). The authors simulated 1-h increments of intensity, translation speed and heading angle. Each dependent variable is condition on one prior. For example, the translation speed at time t_{i+1} is conditional on the translation speed at time t_i, current location and month. The authors derived the PDFs for variable-sized regions centered at every 0.5° latitude and longitude grid cells. However, they did not provide further detail regarding the criteria to select the region size, the interval sizes of the variables during their distribution calibration, or the distribution type (e.g., parametric distribution, non-parametric distribution,

etc.) Similar to Casson and Coles (2000), these authors also did not provide a comparison between the historical and simulated tracks.

James and Mason (2005) developed an autoregression stochastic process for the simulation of the tropical cyclones propagation over the Queensland Coast of Australia. Due to the limited amount of historical data, the authors performed a multiple linear regression using all the data over the Coral Sea without breaking their analysis down to specific regions or a grid of geographic cells. Their model is similar to the model proposed by Vickery et al. (2000a), however, the independent and dependent variables in their analysis were velocity components along the latitude and longitude direction and not velocity magnitude and heading angle as modeled in Vickery et al. (2000a).

Emanuel et al. (2006) introduced two approaches for the storm's track simulation. The first approach is based on Markov Chain theory and uses conditional PDFs to sample the translation acceleration and heading angle rate of change. Many conditional PDFs were constructed and those used are selected based on a sampling schedule in order to ensure that the track simulation also propagates in regions where the historical data are limited. The genesis module in this approach considers time and space dependent PDFs which are developed from histograms of genesis points and smoothing. The second approach introduced by the authors proposes a track generation from synthetic wind time series. The underlying assumption in this technique is that the storm moves with some vertical mean wind which is modeled as an average mean of the winds occurring at two predefined levels plus a drift term. The storm's track simulation methodology developed for this thesis is based on the first approach presented by Emanuel et al. (2006). A more detailed description will follow.

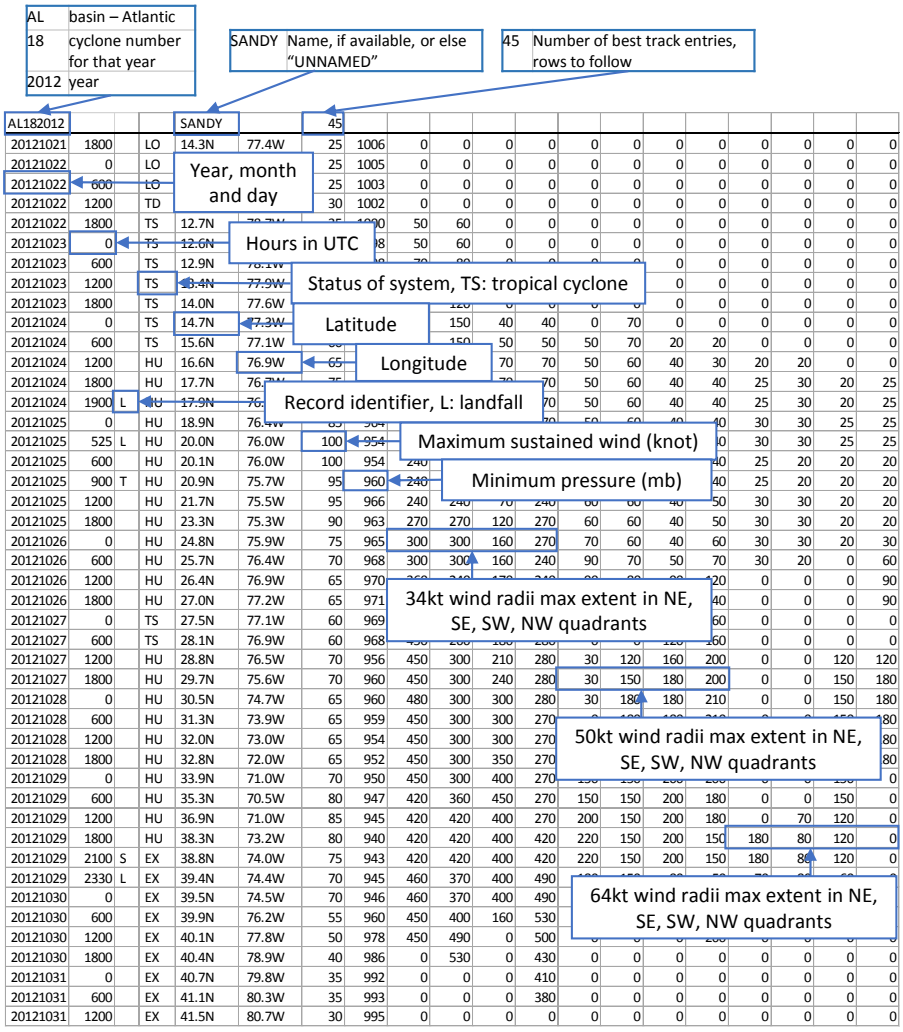
Hall and Jewson (2007) simulated the track propagation by modeling the displacement increments on two perpendicular directions. Their innovation is that at each time step the

authors sample the forward displacement by first assessing all historical increments by inversely weighting with distance from the storm's current location all historical incremental displacements. At each location of the track they compute the following information: i) the weighted mean latitude and longitude by averaging all historical data, ii) the weighted variances parallel and normal to the weighted mean direction, and, iii) the autocorrelation, as a measure of the displacement anomalies, parallel and normal to the mean track again by weighting all historical track data. This approach properly considers the local geography; however, it is computationally expensive. Specifically, for every new location of the storm's track, the algorithm used needs to perform multiple distance calculations, as many as the number of data points in the basin.

5.2.1 Tropical cyclone dataset

The most well-known historical hurricane dataset is the "best-track" database which has been used in all documented research on tropical cyclone track simulation methods. This dataset is the best track and intensity estimates of tropical cyclones as determined in a post-storm analysis of all available data. This database is available for all basins (i.e., areas of tropical cyclone formation). For the North Atlantic basin specifically is documented as the Hurricane Databases (HURDAT) first reported by Jarvinen et al. (1984). The HURDAT dataset has been updated throughout the years and is provided by the National Hurricane Center (NHC) of the National Oceanic and Atmospheric Administration (NOAA). The range of years covered from the HURDAT dataset and used in this work span from 1951 to 2017.

Nowadays, the information to estimate the "best track" is obtained from multiple sources such as satellites, ships and buoys, aerial reconnaissance, radio sondes, radar and automated surface observation systems (ASOS). Historically, however, these data were not accurately recorded. The reliability of the instruments used to record these data increased significantly



AL182012				SANDY		45													
20121021	1800		LO	14.3N	77.4W	25	1006	0	0	0	0	0	0	0	0	0	0	0	0
20121022	0		LO			25	1005	0	0	0	0	0	0	0	0	0	0	0	0
20121022	600		LO			25	1003	0	0	0	0	0	0	0	0	0	0	0	0
20121022	1200		TD			30	1002	0	0	0	0	0	0	0	0	0	0	0	0
20121022	1800		TS	12.7N				50	60	0	0	0	0	0	0	0	0	0	0
20121023	0		TS	12.6N				50	60	0	0	0	0	0	0	0	0	0	0
20121023	600		TS	12.9N										0	0	0	0	0	0
20121023	1200		TS	13.4N	77.9W									0	0	0	0	0	0
20121023	1800		TS	14.0N	77.6W									0	0	0	0	0	0
20121024	0		TS	14.7N				150	40	40	0	70	0	0	0	0	0	0	0
20121024	600		TS	15.6N	77.1W			150	50	50	50	70	20	20	0	0	0	0	0
20121024	1200		HU	16.6N	76.9W	65		70	70	50	60	40	30	20	20	0	0	0	0
20121024	1800		HU	17.7N						70	50	60	40	40	25	30	20	25	
20121024	1900	L	HU	17.9N						70	50	60	40	40	25	30	20	25	
20121025	0		HU	18.9N											40	30	30	25	25
20121025	525	L	HU	20.0N	76.0W	100	954								0	30	30	25	25
20121025	600		HU	20.1N	76.0W	100	954								40	25	20	20	20
20121025	900	T	HU	20.9N	75.7W	95	960								40	25	20	20	20
20121025	1200		HU	21.7N	75.5W	95	966	240							50	30	30	20	20
20121025	1800		HU	23.3N	75.3W	90	963	270	270	120	270	60	60	40	50	30	30	20	20
20121026	0		HU	24.8N	75.9W	75	965	300	300	160	270	70	60	40	60	30	30	20	30
20121026	600		HU	25.7N	76.4W	70	968	300	300	160	240	90	70	50	70	30	20	0	60
20121026	1200		HU	26.4N	76.9W	65	970								120	0	0	0	90
20121026	1800		HU	27.0N	77.2W	65	971								40	0	0	0	90
20121027	0		TS	27.5N	77.1W	60	969								60	0	0	0	0
20121027	600		TS	28.1N	76.9W	60	968								60	0	0	0	0
20121027	1200		HU	28.8N	76.5W	70	956	450	300	210	280	30	120	160	200	0	0	120	120
20121027	1800		HU	29.7N	75.6W	70	960	450	300	240	280	30	150	180	200	0	0	150	180
20121028	0		HU	30.5N	74.7W	65	960	480	300	300	280	30	180	180	210	0	0	150	180
20121028	600		HU	31.3N	73.9W	65	959	450	300	300	270								180
20121028	1200		HU	32.0N	73.0W	65	954	450	300	300	270								80
20121028	1800		HU	32.8N	72.0W	65	952	450	300	350	270								80
20121029	0		HU	33.9N	71.0W	70	950	450	300	400	270								0
20121029	600		HU	35.3N	70.5W	80	947	420	360	450	270	150	150	200	180	0	0	150	0
20121029	1200		HU	36.9N	71.0W	85	945	420	420	400	270	200	150	200	180	0	70	120	0
20121029	1800		HU	38.3N	73.2W	80	940	420	420	400	420	220	150	200	150	180	80	120	0
20121029	2100	S	EX	38.8N	74.0W	75	943	420	420	400	420	220	150	200	150	180	80	120	0
20121029	2330	L	EX	39.4N	74.4W	70	945	460	370	400	490								
20121030	0		EX	39.5N	74.5W	70	946	460	370	400	490								
20121030	600		EX	39.9N	76.2W	55	960	450	400	160	530								
20121030	1200		EX	40.1N	77.8W	50	978	450	490	0	500								
20121030	1800		EX	40.4N	78.9W	40	986	0	530	0	430	0	0	0	0	0	0	0	0
20121031	0		EX	40.7N	79.8W	35	992	0	0	0	410	0	0	0	0	0	0	0	0
20121031	600		EX	41.1N	80.3W	35	993	0	0	0	380	0	0	0	0	0	0	0	0
20121031	1200		EX	41.5N	80.7W	30	995	0	0	0	0	0	0	0	0	0	0	0	0

Figure 5.1: HURDAT2 format for hurricane SANDY from 2012.

during last few decades.

Prior the development of satellites during the 1960s, the data relied heavily on reconnaissance aircraft flying into hurricanes. The first flights collecting valuable data started in 1943 and flying routinely in a annual basis. However, reliable pressure measurements were obtained from planes after the 1940s and it was not until the 1990s when the aircraft reconnaissance would measure directly the winds in a hurricane. The first weather radar network to survey the coastline of the U.S. was established in 1957 by the U.S. Weather Bureau, which is known today as the National Weather Service (NWS) and in the 1990s

the radar network adopted Doppler technology.

Given the technologies reviewed thusfar—which are mostly land-based—one can expect the genesis location of a hurricane to be inaccurate until the introduction of weather satellites in 1960. The first satellite was named TIROS and was launched by NASA in 1960. Soon after, more satellites with a variety of measurement instruments were introduced. Then, in the 2000s microwave imagery greatly improved diagnoses of tropical cyclone strength, intensity, and track location.

The last available revision of the Atlantic hurricane database (HURDAT2) provided by the National Hurricane Center (NHC) is used in this chapter (NHC, 2014). The format of the HURDAT2 data is give in Figure 5.1 through an example of hurricane Sandy from 2012.

5.2.2 Tropical cyclone track simulation

The algorithm developed in this chapter for the tropical cyclone track simulation is based on the work presented by Emanuel et al. (2006). In Emanuel et al. (2006) the authors developed a methodology in which the track's simulation is broken down to three phases: the genesis phase, the transition or track propagation phase and the dissipation phase. For all three phases the authors used the 6-h time step increments of the HURDAT database. They proposed to model the track as a Markov Chain after realizing that the correlation length of the temporal autocorrelation spectra of key motion characteristics of the track such as the heading angle and the translation speed is relatively small. This means that the three 6-h time step priors accurately condition the transition probabilities. Similar to Emanuel et al. (2006) Markov Chain-based methodology, the Markov Chain approach is likewise used in this chapter; however, there are notable differences, which will be discussed in detail in the subsequent paragraphs. This chapter also includes details on implementation and

simulation—details which are omitted in the work published by Emanuel et al. (2006)—thus enhancing this methodology's applicability.

5.2.2.1 Genesis model

The genesis model implemented in this work is a combination of the models proposed by Vickery et al. (2000a) and Emanuel et al. (2006). Emanuel et al. (2006) proposed to construct a three dimensional probability density function of the genesis location and time developed from three-dimensional Gaussian kernel. The smoothing kernel is isotropic with varying scale for the geographic axes and a varying time range (between 5 and 15 days) for the time axis; however, no specific information was provided on the time or space discretization. Vickery et al. (2000a) on the contrary sampled initiation location, date, time, initiation translation speed and heading angle directly from the HURDAT database in an effort to maintain the climatology and the seasonal preference associated with the genesis location.

For the track genesis model developed in this chapter, the post-1970 HURDAT data were utilized to construct a two dimensional PDF from which the new locations are sampled from. The information on the date, time, initiation speed and angle is also adopted from the post-1970 database and specifically from the event that is geographically closest to the simulated genesis location.

The hurricane genesis locations are assumed to occur over water and not over land on the Americas. Therefore, land-based genesis locations are excluded from the subsequent steps. The most recent available data identifies 11 out of 782 hurricane events in the Americas with a land-based genesis location. Having filtered out these data points, the remaining data points are partitioned into square bins of $0.5^\circ \times 0.5^\circ$ which mesh the Atlantic basin, Caribbean Sea and the Gulf of Mexico (Figure 5.2). The selected region to scan for data is

[5N : 55N] in latitude and [-110W : 0W] in longitude. Next, an isotropic Gaussian kernel

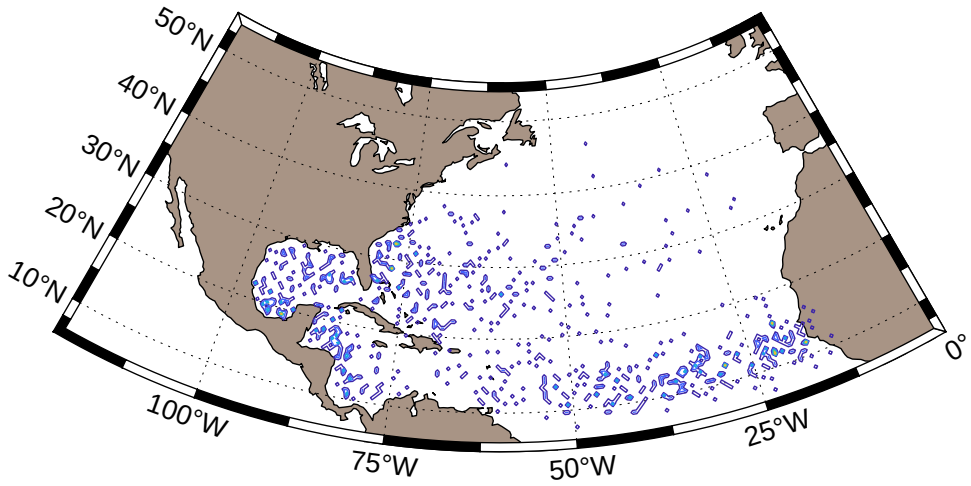


Figure 5.2: Spatial probability distribution of annual genesis locations produced from the post-1970 hurricane data when binned in a $0.5^\circ \times 0.5^\circ$ grid.

with standard deviation of 1° is applied to the binned data. The value of 1° is found to reproduce reasonable results given the amount of data and a kernel with fixed standard deviation. Any overland region which has a nonzero weight is filtered out and the distribution function is renormalized to produce a volume equal to 1. The resulting annual genesis PDF is depicted in Figure 5.3. This approach is considered as a middle-ground approach for the genesis model compared to what was proposed by Vickery et al. (2000a) and Emanuel et al. (2006). The former suggested to sample directly form the historical genesis locations and the latter form a distribution resulting from smoothing with variable but higher (and unknown) than 1° standard deviation at the regions with little activity.

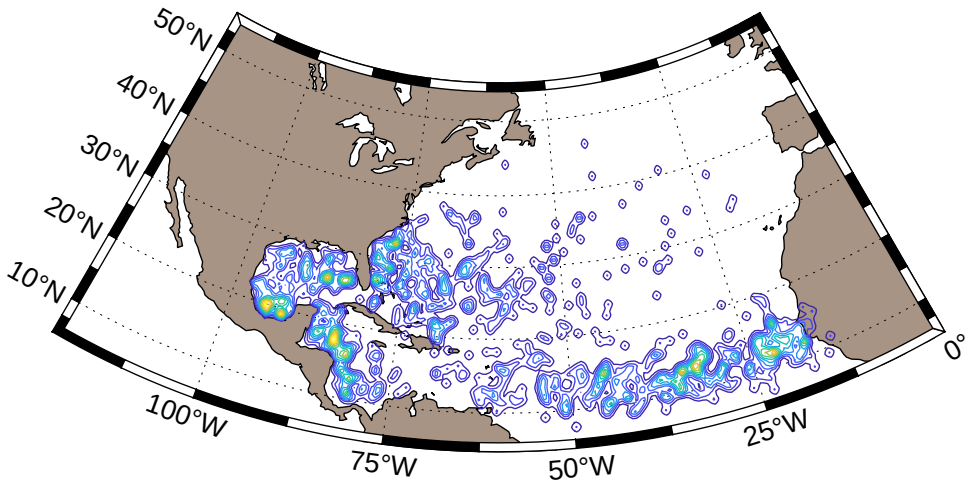


Figure 5.3: Spatial probability distribution of annual genesis locations produced from the post-1970 hurricane data and a two-dimensional isotropic Kernel with $\sigma = 1$.

For the simulation of the genesis locations, a modified version of the inverse cumulative distribution function (CDF) transform method is implemented. We assume that X is a discrete random variable such that $P(X = x_i) = p_i$ and $\sum_{j=1}^{N} P(X = x_j) = \sum_{j=1}^{N} p_j = 1$. Variable X represents the geographic bin where the hurricane genesis occurs. Here it is assumed the the exact location of the hurricane genesis in a $0.5^\circ \times 0.5^\circ$ geographic bin is uniformly distributed. To sample a genesis location, a sample for X must be generated first as the following steps indicate:

- construct discrete CDF from the joint discrete genesis PDF by scanning the bins row-wise. Each bin is assigned an index k and the bin's weight is added to the CDF under construction

- generate $U \sim \text{Unif}(0,1)$
- determine the index k such that $\sum_{j=1}^{k-1} p_j \leq U < \sum_{j=1}^{k} p_j$
- generate $Udlon \sim \text{Unif}(0,1) \cdot 0.5^\circ$ and $Udlat \sim \text{Unif}(0,1) \cdot 0.5^\circ$
- Add $Udlon$ and $Udlat$ to the origin (SE edge) of bin k

After the track's genesis location is defined, the initialization speed and heading angle are sampled. These parameters are sampled from corresponding initiation PDFs which are constructed for each cell of a $2^\circ \times 2^\circ$ geographic grid. For their construction, the first 6-h increments of all post-1970 storm records are utilized. The raw data do not yield robust PDFs for most of the geographic cells and therefore subsequent smoothing operations are performed–one spatial smoothing and one along the variable's range. The spatial smoothing uses data from neighboring geographic cells, by inversely weighting the cells with distance. The data in a neighbor cell are all taking the same weight which is determined based on a two-dimensional isotropic Gaussian function that is centered on the centroid of the cell that the genesis location belongs in. Histogram smoothing is performed in much the same way after spatial smoothing occurs. More details on the implementation of the two smoothing operations are presented in the following section.

The discretization bin of the initialization speed variable and initialization heading angle variable is 20 km/6h and 10 °, respectively, and, they are fixed for all geographic cells. For the spatial smoothing, the data included in 25 geographic cells (5 cells in each direction) are considered and each cell is weighted based on an isotropic two-dimensional Gaussian function with standard deviation equal to 4°. To smooth the histogram of the speed variable, the data included in a window of 7 bins are considered and each bin is weighted based on a unidimensional Gaussian function with standard deviation equal to 20 km/6h. Similarly, to smooth the histogram of the heading angle, the data included in a window of 7 bins are considered and each bin is weighted based on a unidimensional

Gaussian function with standard deviation equal to 10 °. Figure 5.4 displays the smoothed PMFs of initialization translation speed and heading angle for geographic cell with [11N : 13N] latitude and [-52W : -50W] longitude.

To simulate from the empirical PMFs, the inverse CDF transformed method is utilized. One uniform random number is drawn to sample the bin of the empirical PMF, and a second independent uniform (within the range of values of the selected bin) random number is drawn to simulate the actual value. The information of the Coordinated Universal Time (UTC), day and month are adopted from the historical record which genesis point is the closest to the simulated genesis point. In the HURDAT2 database; however, it so happens that more than one storms initiate from the same genesis location. In that case a randomly selected record among those which initiate from the same genesis location is selected to adopt information on time.

5.2.2.2 Propagation model

The track propagation model proposed herein is based on the Markov Chain Theory. At each geographic location, the track transitions from one state to the next based on a transition probability. This probability is expressed as $P(X_i|X_{i-1})$ in its general form, and the outcomes sampled from it determine how the track moves to its next location. The term X is the state variable of the Markov Chain and the index i refers to the i-th 6-hourly increment. The term X_i to the left of the conditional probability is referred to as the predictand random variable and predictor random variable is the term to the right X_{i-1}. The transitional probability considered herein is defined as:

$$P\left(\vec{v}_i|\vec{p}_i,\vec{v}_{i-1}\right) \tag{5.1}$$

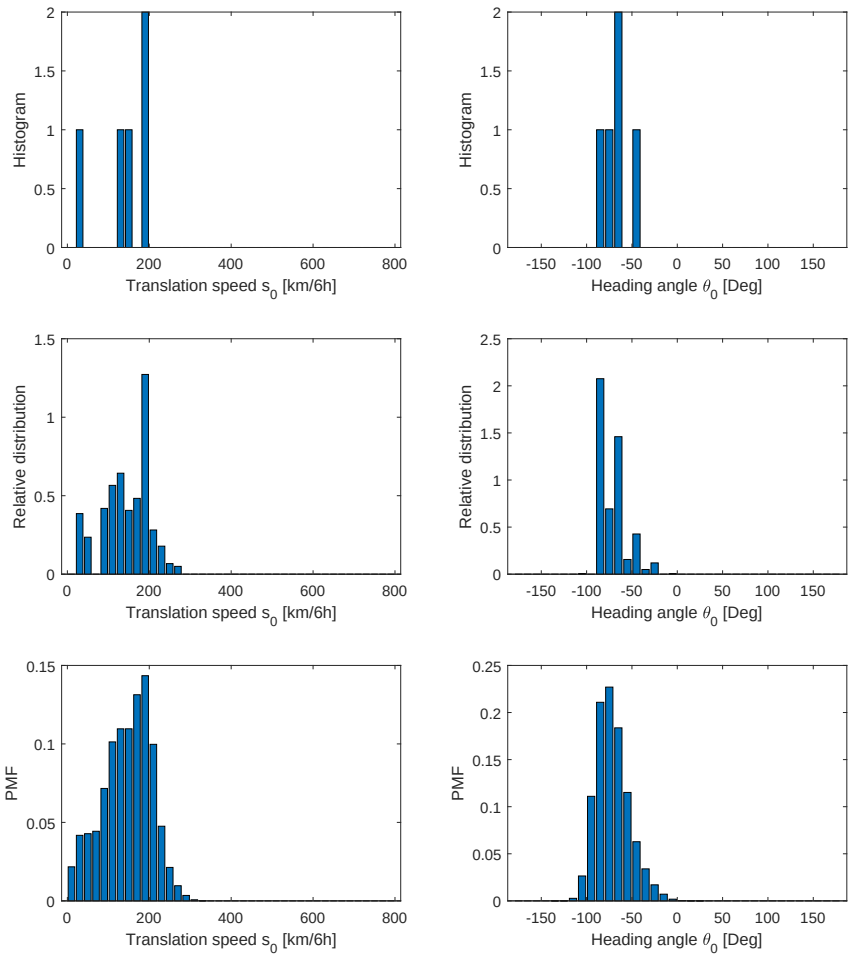


Figure 5.4: Probability mass functions of the initialization translation speed s_0 (left column) and initialization heading angle (right column) for the square geographic cell with centroid coordinates [12N, -51W] and side length equal to 1°. The top row displays the raw data, the middle row displays the spatially smoothed data and the last row shows the final PMFs used in the simulation.

where $\vec{v}_i$ and $\vec{p}_i$ are the random variables of the acceleration vector and position vector respectively at the i-th 6-hourly increment and $\vec{v}_{i-1}$ is the random variable speed vector at the $(i-1)$-th 6-hourly increment. These vectors are defined as follows:

$$\vec{p} = (lon, lat) \tag{5.2}$$

where lon and lat are random variables of the longitude and latitude respectively,

$$\vec{v} = (s, \theta) \tag{5.3}$$

where s and θ are the random variables of the translation speed and heading angle respectively,

$$\vec{\dot{v}} = \frac{d\vec{v}}{dt} = (\dot{s}, \dot{\theta}) \tag{5.4}$$

where $\dot{s}$ and $\dot{\theta}$ are the random variables of the translation speed rate of change (or translation acceleration) and the heading angle rate of change respectively. When these vectors are used with a subscript, this refers to the time step of the track's time history.

Vector $\vec{v}$ is expressed in polar coordinate system and each point on a "plane" is determined by a distance from a reference location and an angle from a reference direction. For any 6-hourly increment i, the reference location is the current geographic position of the track and the reference direction is always a vector facing North. Herein, the "plane" is the surface of the earth and the distance is the great-cycle distance, which by definition is the shortest distance between two points on the surface of a sphere. One can conclude that for equally spaced in time track's translation increments, the velocity magnitude is equivalent to the great-cycle measurement between two locations ($\|\vec{v}\| = s$). Additionally, it is assumed that the track always follows great-cycle paths when moving between consecutive 6-hourly increments.

Replacing the vectors with their components, the probability expressed in (5.1) can be written as:

$$P\left(\dot{s}_i, \dot{\theta}_i | lon_i, lat_i, s_{i-1}, \theta_{i-1}\right) \tag{5.5}$$

This density function is a 6-dimensional PDF, which is not easily quantified. To address its quantification, one should consider the following assumptions:

assump. 1: independence between the rate of change of the translation velocity and the rate of change of the heading angle, that is $\dot{s}_i \perp \dot{\theta}_i$

assump. 2: independence between the rate of change of the translation velocity and the prior heading angle, that is $\dot{s}_i \perp \theta_{i-1}$

assump. 3: independence between the rate of change of the heading angle and the prior translation velocity, that is $\dot{\theta}_i \perp s_{i-1}$

which permit the computation of the transition probability $P\left(\dot{s}_i, \dot{\theta}_i | lon_i, lat_i, s_{i-1}, \theta_{i-1}\right)$ as follows:

$$\begin{aligned} P\left(\dot{s}_i, \dot{\theta}_i | lon_i, lat_i, s_{i-1}, \theta_{i-1}\right) &\overset{\text{ass.1}}{=} P\left(\dot{s}_i | lon_i, lat_i, s_{i-1}, \theta_{i-1}\right) \cdot P\left(\dot{\theta}_i | lon_i, lat_i, s_{i-1}, \theta_{i-1}\right) \\ &\overset{\text{ass.2}}{=} P\left(\dot{s}_i | lon_i, lat_i, s_{i-1}\right) \cdot P\left(\dot{\theta}_i | lon_i, lat_i, s_{i-1}, \theta_{i-1}\right) \\ &\overset{\text{ass.3}}{=} P\left(\dot{s}_i | lon_i, lat_i, s_{i-1}\right) \cdot P\left(\dot{\theta}_i | lon_i, lat_i, \theta_{i-1}\right) \end{aligned} \tag{5.6}$$

Therefore, when a track is located in a geographic cell, it transitions based on the cell's transitional probabilities $P(\dot{s}_i | s_{i-1})$ and $P(\dot{\theta}_i | \theta_{i-1})$. The authors in Emanuel et al. (2006) selected the change rates of the translation speed and heading angle as predictands in order to take advantage of a coarse spatial resolution representation of the transitional probabilities while capturing the smoothness of the track. In Emanuel et al. (2006) the authors pointed out that plotting pairs of historical data of the current speed (s_i) versus the previous states of translation speed (s_{i-1}) results in a nearly linear relationship. In other words, the translation speed varies slowly with time. This makes the rate-based representation beneficial because it makes the transition probability densities insensitive to resolution. Similar conclusions can be drawn for heading angle θ.

Computationally, the translation speed variables, heading angles and their rate of change

are defined by:

- translation speed $s_i = \text{dist}(p_i, p_{i+1})$ in $[rad/6h]$
- heading angle $\theta_i = \text{direc}(p_i, p_{i+1})$ in $[rad]$
- speed rate of change $\dot{s}_i = s_i - s_{i-1}$ in $[rad/6h/6h]$
- angle rate of change $\dot{\theta}_i = \theta_i - \theta_{i-1}$ in $[ran/6h]$

where "dist" is the great cycle distance between two location $p_i (= (lat_i, lon_i))$ and p_{i+1}, "direc" is the angle between the track's velocity vector $\vec{v}_i$ at position i, (that is the "great circle" vector connecting p_i and p_{i+1}) and a vector pointing North. Angles θ are clockwise positive and their range is $[-180°, +180°]$. The great cycle distance is determined in

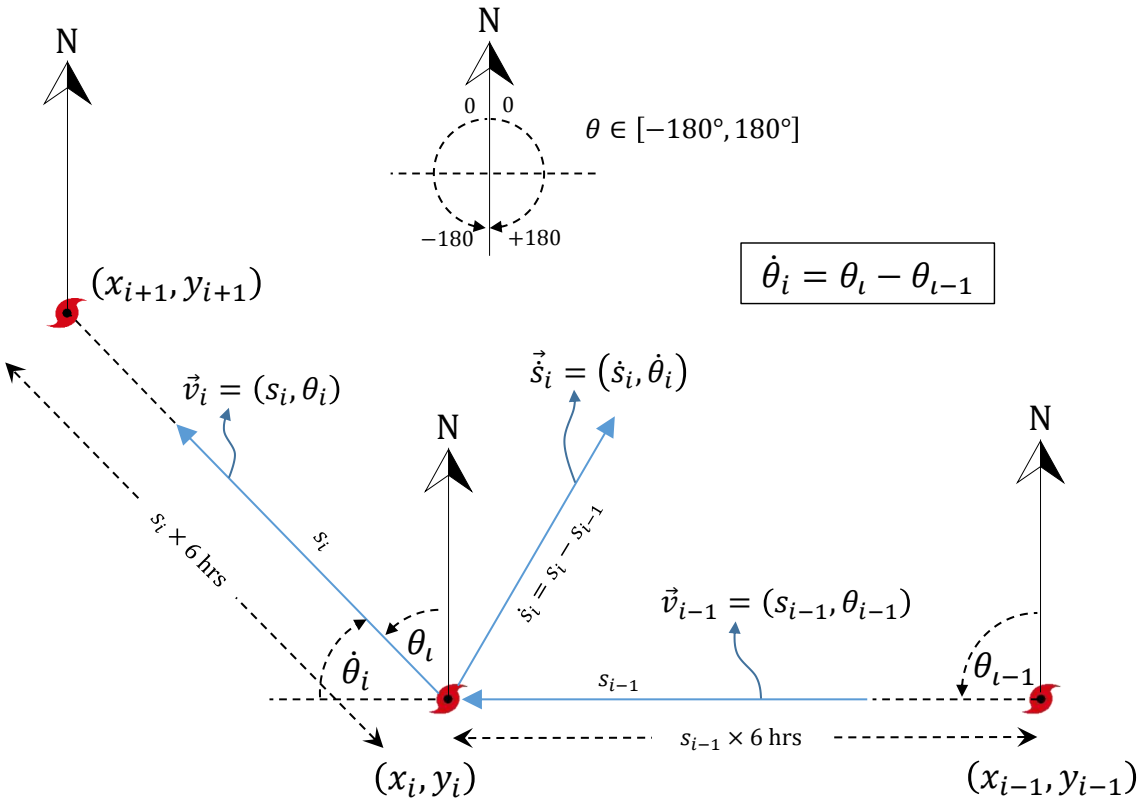


Figure 5.5: A tropical cyclone track crossing three geographical locations in two 6-hourly increments. The blue arrows along the track represent the velocity vectors.

radians by using the haversine formula:

$$\Delta\sigma = 2\arctan\sqrt{\sin^2\left(\frac{\Delta\text{lat}}{2}\right) + \cos\left(\text{lat}_1\right)\cos(\text{lat}_2)\sin^2\left(\frac{\Delta\text{lon}}{2}\right)} \tag{5.7}$$

where $\Delta\sigma$ is the angular distance, Δlon and Δlat are the differences in longitude and latitude, respectively, between two geographic locations. The radial distance can be converted to kilometers by multiplied with the earth radius which is assumed herein to be 6,371 km.

Figure 5.5 shows the path of a tropical cyclone's track moving through three geographical locations in two 6-hourly increments. The corresponding vectors and time parameter indexes of the measurements are displayed. The tranlsation velocity vectors have units of [distance/6h] and the vectors of velocity rate of change have units of [distance/6h/6h].

The values of the translation speed, heading angle and their rate of change are calculated for each tropical cyclone record of the HURDAT database. Figure 5.6 shows corresponding values for each of the 6-hourly increments of Hurricane Sandy from 2012.

The historical data need to be "cleaned" before computing the aforementioned quantities in each track and populate the conditional probabilities. Several data cleaning operations require attention:

- Removal of records with less than three data points from the database. Computations of the speed and direction rate of change require three data points at minimum. In this step 32 records were excluded from the analysis. It should be noted that 31 of the 32 excluded records were recorded prior to 1970.

- Removal of records which do not yield hurricane movement. Three historical storms were found to have multiple static geographic location records and thus excluded from the analysis. These records were also measured prior to 1970.

Sandy		Position		Transl. Speed		Heading Angle	
$Y/M/D$	UTC	$Lat.$	$Lon.$	s_i	$\dot{s}_i$	θ_i	$\dot{\theta}_i$
20121021	1800	14.3N	77.4W	0.00973	N/A	-2.37063	N/A
20121022	0	13.9N	77.8W	0.00973	8.16E-06	-2.36979	0.00084
20121022	600	13.5N	78.2W	0.00974	7.93E-06	-2.36898	0.00081
20121022	1200	13.1N	78.6W	0.00719	-0.00256	-2.90237	-0.53339
20121022	1800	12.7N	78.7W	0.00540	-0.00179	1.89942	-1.48140
20121023	0	12.6N	78.4W	0.00731	0.00192	0.77234	-1.12708
20121023	600	12.9N	78.1W	0.00937	0.00205	0.37104	-0.40130
20121023	1200	13.4N	77.9W	0.01164	0.00228	0.45159	0.08055
20121023	1800	14.0N	77.6W	0.01323	0.00159	0.39289	-0.05869
20121024	0	14.7N	77.3W	0.01607	0.00284	0.21084	-0.18205
20121024	600	15.6N	77.1W	0.01777	0.00171	0.18936	-0.02148
20121024	1200	16.6N	76.9W	0.01949	0.00171	0.17150	-0.01786
20121024	1800	17.7N	76.7W	0.00387	-0.01562	0.44405	0.27255
20121024	1900	17.9N	76.6W	0.01776	0.01390	0.18700	-0.25705
20121025	0	18.9N	76.4W	0.02030	0.00253	0.32917	0.14217
20121025	525	20.0N	76.0W	0.00175	-0.01855	0.00000	-0.32917
20121025	600	20.1N	76.0W	0.01480	0.01305	0.33688	0.33688
20121025	900	20.9N	75.7W	0.01434	-0.00046	0.22821	-0.10867
20121025	1200	21.7N	75.5W	0.02811	0.01377	0.11431	-0.11390
20121025	1800	23.3N	75.3W	0.02787	-0.00024	-0.34810	-0.46241
20121026	0	24.8N	75.9W	0.01758	-0.01029	-0.46377	-0.11567
20121026	600	25.7N	76.4W	0.01452	-0.00306	-0.56862	-0.10485
20121026	1200	26.4N	76.9W	0.01147	-0.00305	-0.41892	0.14970
20121026	1800	27.0N	77.2W	0.00886	-0.00261	0.17557	0.59449
20121027	0	27.5N	77.1W	0.01092	0.00205	0.28592	0.11036
20121027	600	28.1N	76.9W	0.01367	0.00275	0.46392	0.17800
20121027	1200	28.8N	76.5W	0.02085	0.00717	0.71359	0.24967
20121027	1800	29.7N	75.6W	0.01948	-0.00136	0.76794	0.05435
20121028	0	30.5N	74.7W	0.01840	-0.00109	0.70558	-0.06236
20121028	600	31.3N	73.9W	0.01811	-0.00029	0.82637	0.12079
20121028	1200	32.0N	73.0W	0.02030	0.00219	0.80769	-0.01868
20121028	1800	32.8N	72.0W	0.02411	0.00381	0.64470	-0.16299
20121029	0	33.9N	71.0W	0.02547	0.00136	0.28345	-0.36125
20121029	600	35.3N	70.5W	0.02880	0.00333	-0.24477	-0.52822
20121029	1200	36.9N	71.0W	0.03902	0.01022	-0.88247	-0.63770
20121029	1800	38.3N	73.2W	0.01398	-0.02504	-0.89222	-0.00976
20121029	2100	38.8N	74.0W	0.01179	-0.00219	-0.47524	0.41698
20121029	2330	39.4N	74.4W	0.00221	-0.00959	-0.65699	-0.18175
20121030	0	39.5N	74.5W	0.02387	0.02167	-1.26456	-0.60757
20121030	600	39.9N	76.2W	0.02167	-0.00220	-1.40008	-0.13552
20121030	1200	40.1N	77.8W	0.01556	-0.00611	-1.22141	0.17867
20121030	1800	40.4N	78.9W	0.01303	-0.00253	-1.15229	0.06912
20121031	0	40.7N	79.8W	0.00960	-0.00343	-0.75418	0.39811
20121031	600	41.1N	80.3W	0.00873	-0.00087	-0.64201	0.11217
20121031	1200	41.5N	80.7W	N/A	N/A	N/A	N/A

Figure 5.6: Geographic positions of hurricane Sandy in 6-hourly increments extracted from the HURDAT database and the resulting values of the translation speed s_i in [rad/6h], heading angle θ_i in [rad], and their respective rate of change $\dot{s}_i$ in [rad/6h/6h], and $\dot{\theta}_i$ in [rad/6h].

- Removal of starting datapoints from the records until the tropical cyclone develops speed. There are several records indicating storm initiation; however, the storm develops speed at much later time increments. These datapoints can bias the population of the transition probabilities of this geographic cell and can also erroneously yield extreme values of the velocity rate of change and heading angle rate of change. During this cleaning step, 7 records were modified.

- Removal of data from storm records which are not 6-hourly increments. The revised database, HURDAT2 includes additional data points within the 6-hourly increments

in some storms when extra information is available—for example, before landfall or when the storm reaches its maximum intensity. If not removed, these data bias the conditional probabilities.

In total, 38 records were excluded and the subsequent stochastic models used 1,810 of the original 1,848 records provided in the HURDAT database.

The conditional transitional probabilities $P(\dot{s}_i|lon_i, lat_i, s_{i-1})$ and $P(\dot{\theta}_i|lon_i, lat_i, \theta_{i-1})$ constructed in this chapter are kernel smoothed non-parametric distributions of the raw histograms. These transitional probabilities are 4-dimensional probability densities. For estimating them computationally, all predictors and predictands need to be discretized into finite mutually exclusive numerical categories. The Atlantic basin, for example, is discretized into a grid of $1^\circ \times 1^\circ$ cells and its resulting mesh is shown in Figure 5.7. Let

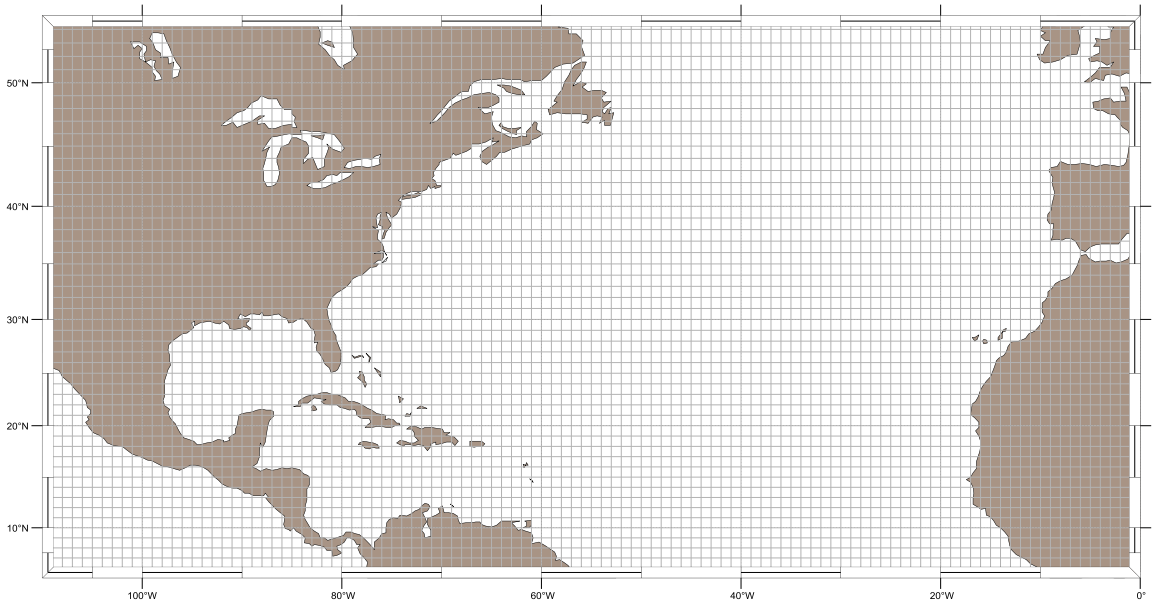


Figure 5.7: Discretization of the Atlantic basin, Gulf of Mexico and Caribbean Sea into a $1^\circ \times 1^\circ$ grid of cells.

position vector (lon_i^k, lat_i^k) refer to the track at time step i when located in geographic cell k. When the transitional probabilities are determined for geographic cell k, these are two-dimensional densities and are expressed as $P\left(\dot{s}_i|lon_i^k, lat_i^k, s_{i-1}\right)$ and $P\left(\dot{\theta}_i|lon_i^k, lat_i^k, \theta_{i-1}\right)$.

When the speed of the track at the prior step is assumed to belong in the p-th speed bin (${s_{i-1}}^p$), then the transitional probability is expressed as $P\left(\dot{s}_i|lon_i^k, lat_i^k, {s_{i-1}}^p\right)$ and is a uni-dimensional probability density function. The probability that the speed rate belongs in the m-th speed rate bin, ($\dot{s}_i^m$), when considering this uni-dimensional probability density, is expressed as $P\left(\dot{s}_i^m|lon_i^k, lat_i^k, {s_{i-1}}^p\right)$ and is a scalar number.

For the development of uni-dimensional probability densities $P\left(\dot{s}_i|lon_i^k, lat_i^k, {s_{i-1}}^p\right)$ and $P\left(\dot{s}_i|lon_i^k, lat_i^k, {s_{i-1}}^p\right)$, a set of steps is presented next. Comments are given only for $P(\dot{s}_i|s_{i-1})$ and the extension to $P(\dot{\theta}_i|\theta_{i-1})$ is straightforward.

Step 1: Convert units of speed to [rad/6h], speed rate to [rad/6h/6h], heading angle to [rad] and heading angle rate to [rad/6h] and parse the pre-processed data to an easy-to-process data array. Each row of this array refers to one data point and includes the following information:

lat_i	lon_i	s_{i-1}	$\dot{s}_i$	θ_{i-1}	$\dot{\theta}_i$

At this stage there are approximately $4.6 \cdot 10^4$ data points.

Step 2: Distribute all data points to the $1^\circ \times 1^\circ$ geographic cells.

Step 3: The data points in each geographic cell are distributed to the preselected speed bins and the data in each speed bin are subsequently distributed to the preselected speed rate bins. The resolution of the speed bins and speed rate bins is [40km/6h] and [8km/6h/6h] respectively. Similarly, the same data are used to populate the preselected heading angle bins and heading angle rate of change bins. Their bin resolution is [20°] and [3°/6h] respectively. The speed range is considered to span 0-700 [km/6h] and the range of the heading angle spans [-180° – 180°]. The range of the speed rate of change and heading angle rate of change are kept variable to keep the array size requirements as low possible. For the considered resolutions, there are

$110 \times 50 \times (700/40) \approx 1 \cdot 10^5$ speed rate histograms that need to be populated. Similarly, there are $110 \times 50 \times (360/20) \approx 1 \cdot 10^5$ histograms of the heading angle rate of change that need to be populated.

Step 4: In order to populate the histograms described in the previous step with the available amount of data, two types of smoothing operations are employed: 1) a spatial smoothing over the geographic region and 2) a smoothing of the empirical distribution of the variable at hand. The spatial smoothing is applied first and the datapoints in each geographic cells are updated by considering the data points from its neighbor cells with some weight. Let the main geographic cell named master cell and the neighbor cells named slave cells. The weights assigned to each cell, including the master and all slaves cells, is determined from the volume prescribed by a two-dimensional isotropic Gaussian distribution centered at the master cell's centroid. The characteristics of the two-dimensional isotropic Gaussian kernel can vary based on the number of data included under the kernel. However, in this chapter a fixed case of the two-dimensional isotropic Gaussian distribution $N(\mu_{2D}, \Sigma_{2D})$ is considered which spans 4° towards each direction from the master cell and thus covers 80 slave cells and 1 master cell.

The distribution parameters are:

$$\mu_{2D} = \begin{bmatrix} \bar{lon} \\ \bar{lat} \end{bmatrix}, \quad \Sigma_{2D} = \begin{bmatrix} \sigma_{lon}^2 & 0 \\ 0 & \sigma_{lat}^2 \end{bmatrix}$$

The weight of each slave geographic cell is determined as follows:

$$\begin{aligned} w_{\text{cell}} = & \frac{1}{2\pi\sigma_{lon}\sigma_{lat}} \int_{\text{lon}_{\text{West}}}^{\text{lon}_{\text{East}}} \int_{\text{lat}_{\text{South}}}^{\text{lat}_{\text{North}}} \cdot \\ & \exp\left(-\frac{1}{2} \left[\frac{\left(lon - \bar{lon}\right)^2}{\sigma_{lon}^2} + \frac{\left(lat - \bar{lat}\right)^2}{\sigma_{lat}^2} \right] \right) \mathrm{d}lat \; \mathrm{d}lon \end{aligned} \tag{5.8}$$

where lon_{West}, lon_{East}, $\text{lat}_{\text{South}}$ and $\text{lat}_{\text{North}}$ are the geographic boundaries of cell (k,l), where k and l are the relative indices along the longitude and latitude respectively, parameters $\bar{lat}$ and $\bar{lon}$ are the latitude and longitude coordinates of the master's cell centroid. The values of the standard deviations considered herein are $\sigma_{lat} = 2^{\circ}$ and $\sigma_{lon} = 2^{\circ}$. Figure 5.8 displays the weight of each cell when the described 2D Gaussian distribution is considered.

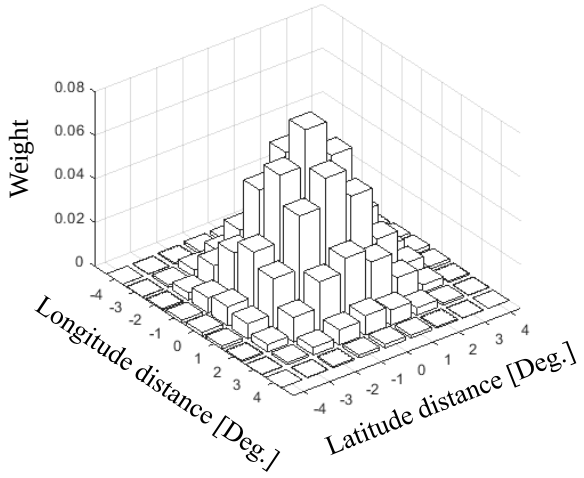


Figure 5.8: Weights assigned to each cell during the spatial smoothing resulting form a two-dimensional Gaussian kernel.

After the spatial smoothing of a geographic cell is performed, the speed rate of change conditional to a prior speed bin is a relative distribution which can be smoothed even further. A uni-dimensional Gaussian kernel is applied to each of the speed rate bins. When the kernel is centered at a bin (master), this bin's information is updated using the information from the neighbor bins with some weight. The weight of each bin is determined from the area under the Gaussian kernel $N(\mu_{1D}, \sigma_{1D})$, which is computed as follows:

$$w_{\text{bin}} = \frac{1}{2\pi\sigma_{1D}} \int_{\text{bin}_{\text{LowerBoundary}}}^{\text{bin}_{\text{UpperBoundary}}} \exp\left(-\frac{(x-\mu_{1D})^2}{2\sigma_{1D}^2}\right) \mathrm{d}x \tag{5.9}$$

where $\text{bin}_{LowerBoundary}$ and $\text{bin}_{UpperBoundary}$ are the lowest and highest values respec-

tively of the bin's range, and parameter μ_{1D} is the master's bin centroid. The characteristics of the uni-dimensional Gaussian kernel can vary; however, in this chapter the parameters are fixed and the kernel covers 3 slave bins to the left and right of the master bin. When the kernel is used to smooth the values of the speed rate of change, a standard deviation $\sigma_{1D} = 8$ [km/6h] is used, whereas, when it is used to smooth the values of the heading angle rate of change, a standard deviation $\sigma_{1D} = 3^{\circ}$ is considered. Figures 5.9 and 5.10 display the weight of the bins when a Gaussian kernel is used for the speed rate bins and the heading angle rate bins respectively.

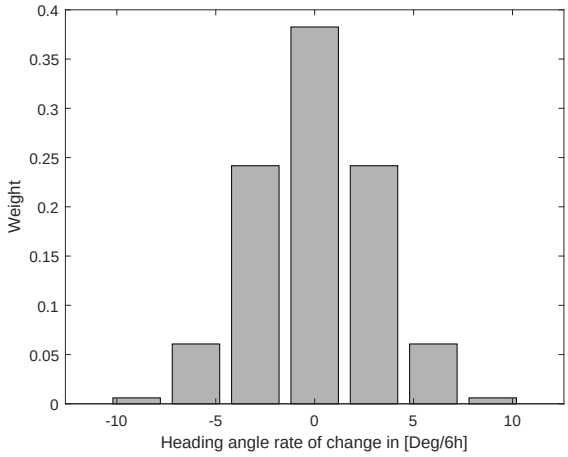


Figure 5.9: Weights assigned to each bin of the heading angle rate of change resulting from a uni-dimensional Gaussian kernel.

Summarizing the steps mentioned above, the probability of a speed rate belong in speed rate bin m while the track is located in geographic cell $(\bar{k},\bar{l})$ with a speed at its prior step which belongs in speed bin p is determined by:

$$P(\dot{s}^m|lon^{\bar{k}}, lat^{\bar{l}}, s^p) = \\ = \frac{\sum_{p=\bar{p}-3}^{p=\bar{p}+3} {w_{\text{bin}}}^p \cdot \left[\sum_{l=\bar{l}-4}^{l=\bar{l}+4}\sum_{k=\bar{k}-4}^{k=\bar{k}+4} {w_{\text{cell}}}^{k,l} H(\dot{s}^m, lon^k, lat^l, s^p)\right]}{\sum_m \left(\sum_{p=\bar{p}-3}^{p=\bar{p}+3} {w_{\text{bin}}}^p \cdot \left[\sum_{l=\bar{l}-4}^{l=\bar{l}+4}\sum_{k=\bar{k}-4}^{k=\bar{k}+4} {w_{\text{cell}}}^{k,l} H(\dot{s}^m, lon^k, lat^l, s^p)\right]\right)} \tag{5.10}$$

where $H(\dot{s}^m, lon^k, lat^l, s^p)$ is the number of data points falling in the geographic cell $(k.l)$,

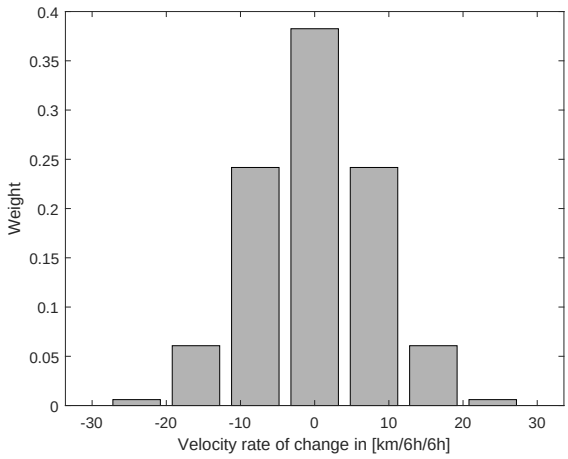


Figure 5.10: Weights assigned to each bin of the heading angle rate of change resulting from a uni-dimensional Gaussian kernel.

with speed at the prior step which falls in bin p and speed rate that belongs in speed rate bin m.

To propagate the track from one location to the next, the speed and heading angle are required. For a given location of the track, prior speed and heading angle, the track is propagating by first sampling probabilities $P(\dot{s}_i|lon_i, lat_i, s_{i-1})$ and $P(\dot{\theta}_i|lon_i, lat_i, \theta_{i-1})$ to set the speed rate of change bin and angle rate of change bin, respectively. After the bins are sampled, a random number is drawn in each case from a uniform distribution with its range defined from the bin's edges. The speed and heading angle are calculated by using the sampled speed rate of change and heading angle rate of change, given by:

$$\begin{aligned} s_i &= s_{i-1} + \dot{s}_i \cdot (6h) \\ \theta_i &= \theta_{i-1} + \dot{\theta}_i \cdot (6h) \end{aligned} \tag{5.11}$$

Recalling that when the speed increments are equally spaced in time, they are synonymous with measurements of displacement. Therefore, Equation (5.11) sets the distance and head-

ing angle for the subsequent position, which is determined by:

$$\begin{aligned} \text{lat}_{i+1} &= \text{asin}\left[\sin(\text{lat}_i)\cdot\cos(\delta)+\cos(\text{lat}_i)\cdot\sin(\delta)\cdot(\theta_i)\right] \\ \text{lon}_{i+1} &= \text{lon}_i+\text{atan2}\left[\sin(\theta_i)\cdot\sin(\delta)\cdot\cos(\text{lat}_i),\cos(\delta)-\sin(\text{lat}_i)\cdot\sin(\text{lat}_i)\right] \end{aligned} \tag{5.12}$$

where $(\text{lat}_i, \text{lon}_i)$ are the latitude and longitude of the current location, δ is the angular distance covered to reach forward location $i+1$ and is defined as $\delta = s_i/R$ where R is the earth's radius equal to 6,371 km.

5.2.2.3 Termination model

The track is terminated based on two criteria: One, the point at which the track crosses the geographic termination boundary, and two, when the track's maximum intensity becomes lower than a predefined threshold. The predefined threshold is expressed here in terms of central pressure and is set to mb greater than 1,012. The termination boundary is a free-form of the area with intense hurricane activity. Termination occurs following the steps below:

Step 1: Collection of tropical cyclone activity data for a $2^\circ \times 2^\circ$ grid mesh of the Atlantic basin

Step 2: Application of a two-dimensional Gaussian isotropic kernel with standard deviation equal to $^\circ$

Step 3: Linear interpolation of the smoothed activity to obtain a $0.5^\circ \times 0.5^\circ$ resolution mesh

Step 4: Retainment of bins with smoothed activity exceeding the value of 12.5

The resulting termination boundary is displayed in Figure 5.11. The second termination criterion—the track's maximum intensity crossing a predefined threshold—is applied after

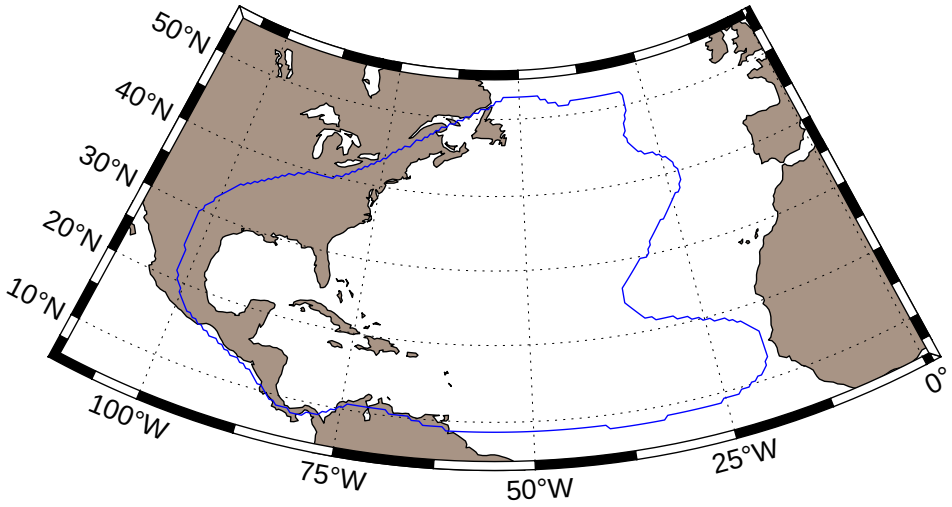


Figure 5.11: Termination boundary of the tropical cyclone activity.

the track and its intensity have been generated. For the track's intensity, please refer to Section 5.3.1. In addition to the two termination criteria described above, the track can also be terminated when it requires conditional PDFs that are unavailable for a given position and value of the prior speed and/or heading angle. To avoid such a scenario, Emanuel et al. (2006) proposed a sampling scheduler in which the geographic cell size increases gradually, first from $0.5^\circ \times 0.5^\circ$ to $5^\circ \times 5^\circ$, then from three mutually exclusive latitude belts, and finally to a single global bin. In this framework, when the track fails to propagate due to lack of data, a second level of conditional distribution is utilized which changes only the parameters of the two-dimensional Gaussian kernel while keeping everything else fixed. In this case, the two-dimensional isotropic Gaussian kernel spans 8° towards each direction with standard deviation $\sigma_{lat} = 4^\circ$ and $\sigma_{lon} = 4^\circ$. If the track continues to fail when

the wider spatial window is used, then the analysis terminates and the track is sampled from the beginning.

5.2.2.4 Test of independence

This section evaluates the assumptions of independence presented in Equation (5.6) by utilizing a chi-square test of independence. The three independence assumptions introduced in Equation (5.6) are tested. In all of them, the population used for analysis is the collection of data enclosed in a geographic cell. Two types of spatial discretizations are used–one with a fine mesh and one with a coarser mesh. The study region in Figure 5.12 is discretized into

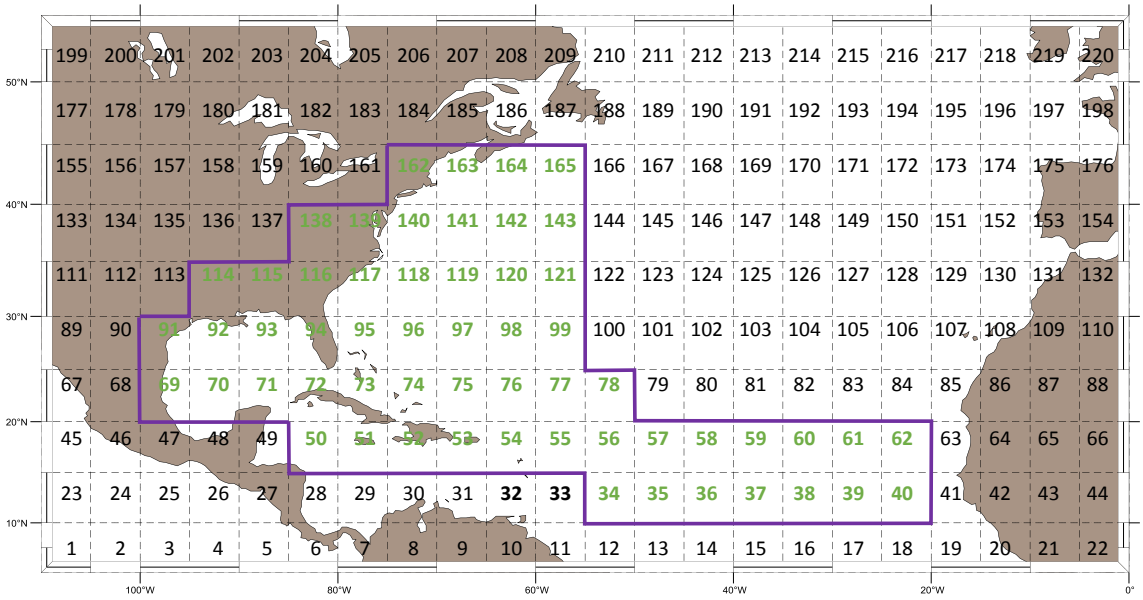


Figure 5.12: Discretization of the Atlantic basin, Gulf of Mexico and Caribbean Sea into a $5^\circ \times 5^\circ$ grid of cells. The cells which are subjected to the χ^2 test of independence are enclosed into the magenta polygon and are enumerated in green color.

$5^\circ \times 5^\circ$ grid cells and the investigated cells are included in the purple polygon with corresponding green reference numbers. Similarly, in Figure 5.13 the study region is discretized into a grid of $10^\circ \times 20^\circ$ cells and the cells being examined are in green. The $5^\circ \times 5^\circ$ discretization is selected because this is the size of the cells used to calibrate the transitional probabilities. However, because of the limited number of data points in many cells of the

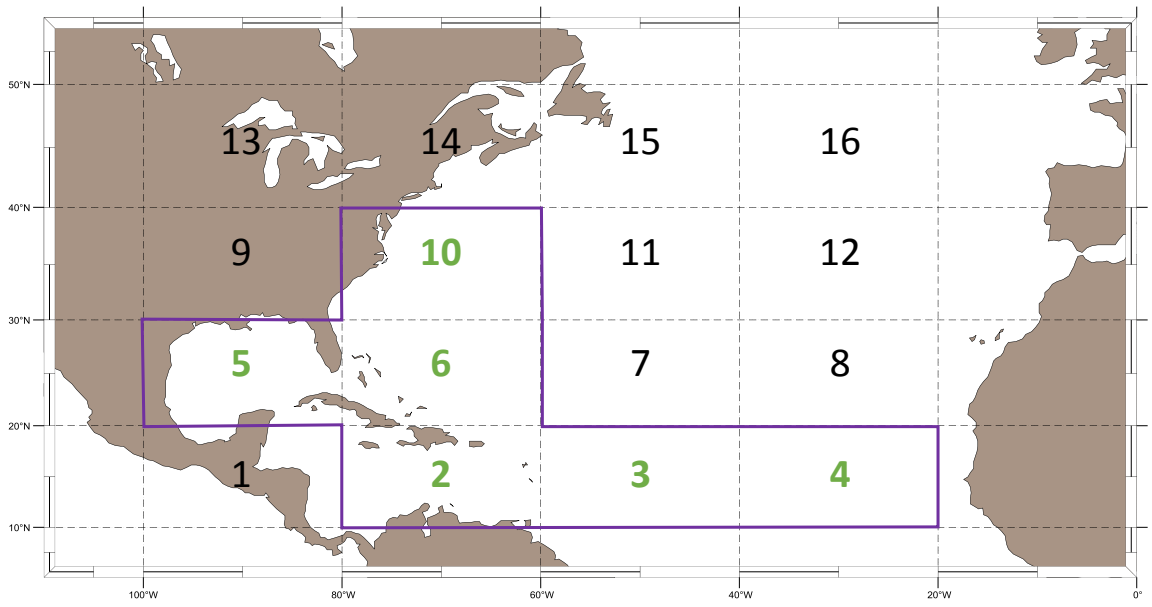


Figure 5.13: Discretization of the Atlantic basin, Gulf of Mexico and Caribbean Sea into a $10^\circ \times 20^\circ$ grid of cells. The cells which are subjected to the χ^2 test of independence are enclosed into the magenta polygon and are enumerated in green color.

refined mesh, a more coarse discretization is examined as well. In this case only the data from post-1970 tropical cyclone records are used. The two hypotheses—common in all tests—are:

H_0 : The investigated variables are independent

H_1 : The investigated variables are not independent

The chi-square test statistic and its p-value are computed under the assumption that H_0 is true —that is that the variables are independent.

Assumption 1 is investigated for 57 of the $5^\circ \times 5^\circ$ size cells and the results are depicted in Figure 5.14. Out of the 57 cells, 23 result in a p-value which exceeds 0.005 and only 14 of them exceed 0.05. The null hypothesis is rejected in all 6 coarse-mesh cells. The results of the χ^2 test for assumption 2 are depicted in Figure 5.15. In this case, 43 cells result in a p-value which exceeds 0.005 and 40 of them exceed 0.05. For the case of $10^\circ \times 20^\circ$

Assumption 1 - Grid (5 x 5)				
Cell	chi2 statistic	*p*-value	dof	ndata
34	383.3310442	3.06E-14	196	328
35	284.380	3.74E-05	196	382
36	399.347	5.20E-16	196	389
37	405.154	1.14E-16	196	411
38	350.500	7.68E-11	196	364
39	232.687	6.61E-03	182	354
40	316.507	1.07E-07	196	225
50	378.072	1.13E-13	196	1110
51	274.686	1.77E-04	196	642
52	279.695	8.02E-05	196	485
53	331.122	5.35E-09	196	545
54	361.818	5.63E-12	196	783
55	442.081	4.81E-21	196	632
56	446.488	1.38E-21	196	568
57	366.670	1.79E-12	196	544
58	259.790	0.000135007	182	433
59	318.363	7.36E-08	196	327
60	321.985	3.55E-08	196	272
61	227.592	1.22E-02	182	176
62	248.681	6.41E-05	169	136
69	286.748	2.52E-05	196	585
70	305.920	8.24E-07	196	876
71	327.792	1.07E-08	196	955
72	293.527	7.85E-06	196	964
73	259.477	1.61E-03	196	651
74	251.596	4.50E-03	196	808
75	354.563	3.04E-11	196	752
76	351.269	6.45E-11	196	636
77	249.115	6.11E-03	196	489
78	230.617	4.57E-02	196	395
91	248.821	0.006333529	196	590
92	297.690	3.75E-06	196	1118
93	267.000	5.59E-04	196	1115
94	242.428	0.013384346	196	1151
95	285.594	3.05E-05	196	1145
96	294.208	6.96E-06	196	881
97	231.708	0.041144074	196	922
98	224.928	0.076634012	196	684
99	289.166	1.67E-05	196	443
114	190.314	6.01E-01	196	328
115	240.765	1.61E-02	196	600
116	226.597	0.066181809	196	658
117	284.709	3.54E-05	196	1340
118	294.642	6.45E-06	196	1073
119	208.923	0.250670016	196	746
120	212.305	0.201798797	196	612
121	206.240	0.293918744	196	502
138	183.019	0.464844137	182	105
139	201.117	0.385969434	196	389
140	224.539	0.079243799	196	742
141	242.710	1.30E-02	196	694
142	187.808	0.650212664	196	556
143	226.132	0.068969526	196	433
162	305.910	2.44E-08	182	190
163	174.620	0.861669153	196	279
164	217.872	0.135781068	196	364
165	188.665	0.633608097	196	310

Figure 5.14: Chi-square test of independence of assumption 1 for the $5^\circ \times 5^\circ$ grid.

cells, 5 out of 6 cells exceed 0.005 and 4 of them exceed 0.05. Lastly, for assumption 3, and in the case of the fine mesh 37 out of 57 cells result in a *p*-value which exceeds 0.005 and 33 of them exceed 0.05. These results conclude that assumption 1 is quite strong as the outcome for most of the cells reject the null hypothesis of independency between the rate of change of the translation speed and the rate of change of the heading angle. Assumptions 2 and 3 are more reasonable because most of the investigated cells do not reject the null hypothesis of independence at the 0.05 level of significance. In other words, assuming that the speed rate of change being independent of the prior heading angle and the rate of change of the heading angle being independent of the prior translating speed are

Assumption 2 - Grid (5 x 5)				
Cell	chi2 statistic	*p*-value	dof	#of_datapoints
34	82.216	0.9511	105	328
35	495.329	0.0000	147	382
36	319.072	0.0000	119	389
37	80.296	0.9651	105	411
38	119.129	0.9831	154	364
39	196.842	0.0112	154	354
40	223.017	0.0000	114	225
50	530.343	0.1187	493	1110
51	559.665	0.0000	416	642
52	474.511	0.6839	490	485
53	327.486	0.0460	286	545
54	313.903	1.0000	480	783
55	272.067	0.1958	253	632
56	638.797	0.0000	200	568
57	369.974	0.0000	266	544
58	180.205	0.1556	162	433
59	481.835	0.0000	161	327
60	124.476	1.0000	209	272
61	123.239	0.9973	170	176
62	132.312	0.3325	126	136
69	371.936	1.0000	512	585
70	576.630	0.0000	442	876
71	463.267	0.9318	510	955
72	553.159	0.3835	544	964
73	396.534	0.9905	465	651
74	395.256	0.9652	448	808
75	730.586	0.0000	390	752
76	249.595	0.9411	286	636
77	591.506	0.0000	320	489
78	570.489	0.0000	322	395
91	727.168	0.0002	595	590
92	597.438	0.9675	663	1118
93	567.749	0.9478	624	1115
94	772.938	1.0000	952	1151
95	637.453	0.2307	612	1145
96	584.591	0.0122	510	881
97	657.446	0.0001	527	922
98	612.483	0.0000	448	684
99	299.240	0.9809	352	443
114	337.576	0.6023	345	328
115	508.449	0.9999	630	600
116	575.229	1.0000	752	658
117	689.340	1.0000	901	1340
118	566.732	0.6232	578	1073
119	457.959	0.9523	510	746
120	688.234	0.0506	629	612
121	429.976	0.9957	510	502
138	213.137	0.9848	260	105
139	600.292	1.0000	765	389
140	640.318	0.9774	714	742
141	389.632	1.0000	629	694
142	415.415	1.0000	612	556
143	474.810	0.9994	578	433
162	265.286	1.0000	440	190
163	460.137	0.2447	440	279
164	488.099	0.3892	480	364
165	422.897	0.9908	494	310

Figure 5.15: Chi-square test of independence of assumption 2 for the $5^\circ \times 5^\circ$ grid

acceptable approximations. Therefore, some deviations of the model from the historical data are expected to occur due to these simplifications.

Assumption 3 - Grid (5 x 5)				
Cell	chi2 statistic	*p*-value	dof	#of_datapoints
34	291.722	1.31E-09	161	328
35	444.549	1.17E-31	147	382
36	326.112	6.91E-17	140	389
37	133.731	0.776147	147	411
38	94.304	0.910414	114	364
39	131.097	0.000989	85	354
40	445.125	1.75E-33	140	225
50	344.403	4.69E-01	343	1110
51	329.320	3.48E-01	320	642
52	256.616	0.220103	240	485
53	281.933	5.59E-05	196	545
54	320.240	6.34E-06	217	783
55	351.505	6.99E-12	189	632
56	396.259	1.05E-11	224	568
57	462.054	5.45E-13	264	544
58	325.749	3.62E-11	175	433
59	150.355	0.000847	100	327
60	205.675	1.10E-01	182	272
61	163.475	0.004292	119	176
62	74.162	0.944046	95	136
69	230.703	0.860442	255	585
70	348.262	9.15E-04	270	876
71	417.582	3.10E-05	308	955
72	308.647	0.855274	336	964
73	253.228	0.466408	252	651
74	254.793	9.52E-01	294	808
75	305.636	5.07E-03	245	752
76	228.646	0.179699	210	636
77	205.369	0.948769	240	489
78	404.208	2.01E-14	210	395
91	1437.663	3.24E-125	371	590
92	290.929	7.99E-01	312	1118
93	335.545	7.02E-01	350	1115
94	409.031	1.00E+00	558	1151
95	617.583	1.01E-06	459	1145
96	418.439	9.75E-01	477	881
97	692.678	2.66E-06	531	922
98	479.022	2.50E-01	459	684
99	338.146	9.83E-03	280	443
114	237.570	0.894407	266	328
115	308.073	0.674	320	600
116	296.399	0.999791	387	658
117	630.663	9.56E-01	693	1340
118	666.839	6.74E-01	684	1073
119	484.608	9.82E-01	552	746
120	535.363	0.003415	450	612
121	467.432	9.73E-01	528	502
138	258.358	0.011343	209	105
139	361.828	0.796244	385	389
140	533.611	2.78E-02	473	742
141	583.931	0.652005	598	694
142	592.040	5.84E-01	600	556
143	552.850	5.65E-01	559	433
162	214.548	0.729526	228	190
163	285.526	0.944018	325	279
164	341.803	0.989939	405	364
165	501.132	0.113217	464	310

Figure 5.16: Chi-square test of independence of assumption 3 for the $5^\circ \times 5^\circ$ grid

5.3 Tropical cyclone intensity

The widely acceptable indicator of the tropical cyclone's intensity is its central pressure at the storm's eye. In general, larger central pressure difference causes larger pressure gradient forces and thus stronger winds. However, in the stochastic models, central pressure is not a desirable variable from which to base a simulation, as its values must fall within certain limits to have physical meaning. Darling (1991) introduced a relative intensity measure which transforms the central pressure to a non-dimensional index I, and first used

Assumption 1 - Grid (10 x 20)				
Cell	chi2 statistic	*p*-value	dof	ndata
2.00E+00	8.39E+02	3.97E-12	576	1.37E+03
3.00E+00	1.12E+03	1.65E-37	576	1.97E+03
4.00E+00	1.05E+03	3.15E-30	576	1.38E+03
5.00E+00	7.75E+02	5.45E-08	576	2.49E+03
6.00E+00	6.99E+02	0.000331	576	2.17E+03
10	7.23E+02	2.76E-05	576	2.32E+03

Assumption 2 - Grid (10 x 20)				
Cell	chi2 statistic	*p*-value	dof	ndata
2	352.91	0.30341	340	1366
3	285.49	0.71733	300	1966
4	297.30	0.22828	280	1382
5	410.77	0.00507	340	2492
6	418.05	0.00243	340	2173
10	380.15	0.06576	340	2324

Assumption 3 - Grid (10 x 20)				
Cell	chi2 statistic	*p*-value	dof	ndata
2	570.082	7.14E-02	522	1366
3	887.265	5.34E-48	352	1966
4	2003.487	1.98E-232	342	1382
5	769.747	6.57E-02	712	2492
6	953.989	8.79E-15	640	2173
10	1097.423	3.45E-02	1014	2324

Figure 5.17: Chi-square test of independence all three assumptions for the $10^\circ \times 10^\circ$ grid

it in single-point simulation. This measure takes into account the concept of the hurricane as a heat engine which is constrained by the inflow (sea surface) and outflow (tropopause surface) temperatures.

5.3.1 Relative intensity model

The tropical cyclone's intensity is estimated by the changes in the central pressure as related to this relative intensity index. Vickery et al. (2000a) used first this index to construct a model for the storm's intensity evolution along its path. In this model, the track's relative intensity at the forward location is modeled as a multiple linear regression coupled to the sea surface temperature. When the storm's eye is located over the ocean, the values of I are

obtained from:

$$\ln(I_{i+1}) = c_0 + c_1 \ln(I_i) + c_2 \ln(I_{i-1}) + c_3 \ln(I_{i-2}) + c_4\, SST_{i+1} + c_5\, (SST_{i+1} - SST_i) + \varepsilon_I \tag{5.13}$$

where I_i is the relative intensity at the i-th time increment, SST_i is the sea surface temperature at the location of the storm's eye at time increment i, and parameters $c_0 - c_5$ are the coefficients of the input variables.

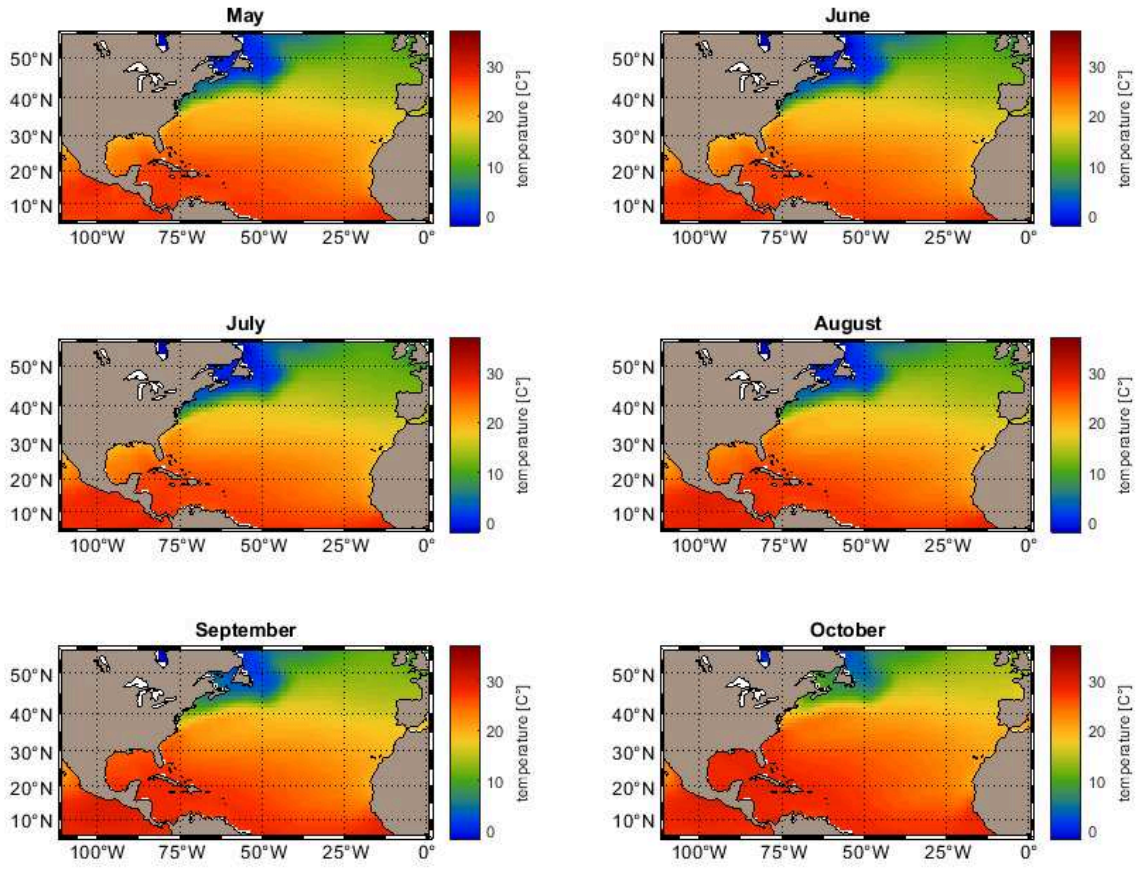


Figure 5.18: Monthly average sea surface temperature (SST) in Celsius over the years of available data provided in the HadISST database. Mean variations of the STT over the Atlantic basin, Gulf of Mexico and Caribbean Sea are in display for the months May to October.

The procedure to derive these coefficients follows the guidelines suggested by Vickery et al. (2000a). The coefficients $c_0, \ldots, c_5$ of the regression model vary with geographic location and direction among other parameters. The coefficients are estimated using a

grid of $5^\circ \times 5^\circ$ cells over the Atlantic basin and a different set is assessed for easterly and westerly traveling hurricanes. Geographic cells without the required number of data match the derived coefficients of the closest cell with adequate number of data.

The sea surface temperature (SST) at the storm center is determined based on the HadISST (Hadley Centre Global Sea Ice and Sea Surface Temperature) dataset, obtained from www.metoffice.gov.uk. The dataset provides monthly mean SSTs for the period 1870-2018 on a $1^\circ \times 1^\circ$ geographical grid in degrees Celsius, which was averaged throughout the years from each month. Figure 5.18 displays the average SSTs for the months between May and October.

The relative intensity index I introduced by Darling (1991) utilizes a simplified version of the minimum central pressure formula—introduced by Emanuel (1988a)— defining a hurricane's intensity which models the tropical cyclone as a heat engine (Emanuel, 1988b). For the calculation of the relative intensity the following definitions and steps are followed. The notation has been adopted from Darling (1991). The surface value of the partial pressure of ambient dry air, p_{da} in mb is given by:

$$p_{da} = 1013 - (\mathrm{RH} \times e_s) \tag{5.14}$$

where RH indicates the relative humidity of ambient air given by RH$= 0.75$, e_s indicates the saturation vapor pressure, given by:

$$e_s = 6.112 \exp\left\{\frac{17.67(T_s - 273)}{t_s - 29.5}\right\} \tag{5.15}$$

where T_s and T_0 indicate the temperatures in degrees kelvin of the ocean surface and of the top of the troposphere (which is assumed to be at the 100-mb pressure) at the center of the tropical cyclone.

The minimum sustainable surface value of central pressure (of dry air) for a hurricane is p_{dc} give by:

$$p_{dc} = x\, p_{da} \tag{5.16}$$

where x is the solution of the following nonlinear equation:

$$\ln(x) = -A\left[\frac{1}{x} - B\right] \tag{5.17}$$

where parameter A is defined by:

$$A \equiv \frac{\varepsilon\, L_v\, e_s}{(1-\varepsilon)\, R_v\, T_s\, p_{da}} \tag{5.18}$$

parameter B is defined by:

$$B \equiv \mathrm{RH}\left[1 + \frac{e_s\, \ln(\mathrm{RH})}{p_{da}\, A}\right] \tag{5.19}$$

where R_v indicates the gas constant of water vapor, given by $R_v = 461$, ε indicates the efficiency of the cyclone as a heat engine and given by:

$$\varepsilon = \frac{T_s - T_0}{T_s} \tag{5.20}$$

parameter L_v indicates the latent heat of vaporization given by:

$$L_v = 2.5\ (10^6) - 2320\ (T_s - 273) \tag{5.21}$$

Nonlinear Equation 5.17 can be solved using different solvers (Newton-Raphson, Secant, fixed point), however, for the results obtained in this chapter a bisection method was implemented.

The relative intensity I is given by:

$$I \equiv \frac{p_{da} - (\tilde{p} - e_s)}{p_{da} - p_{dc}} \tag{5.22}$$

where $\tilde{p}$ is supposed to be the actual central pressure of a tropical cyclone under the prescribed climatic conditions, and the term in the parenthesis $(\tilde{p} - e_s)$ named $\tilde{p}_d$ in Darling (1991) stands for the corresponding partial pressure for dry air.

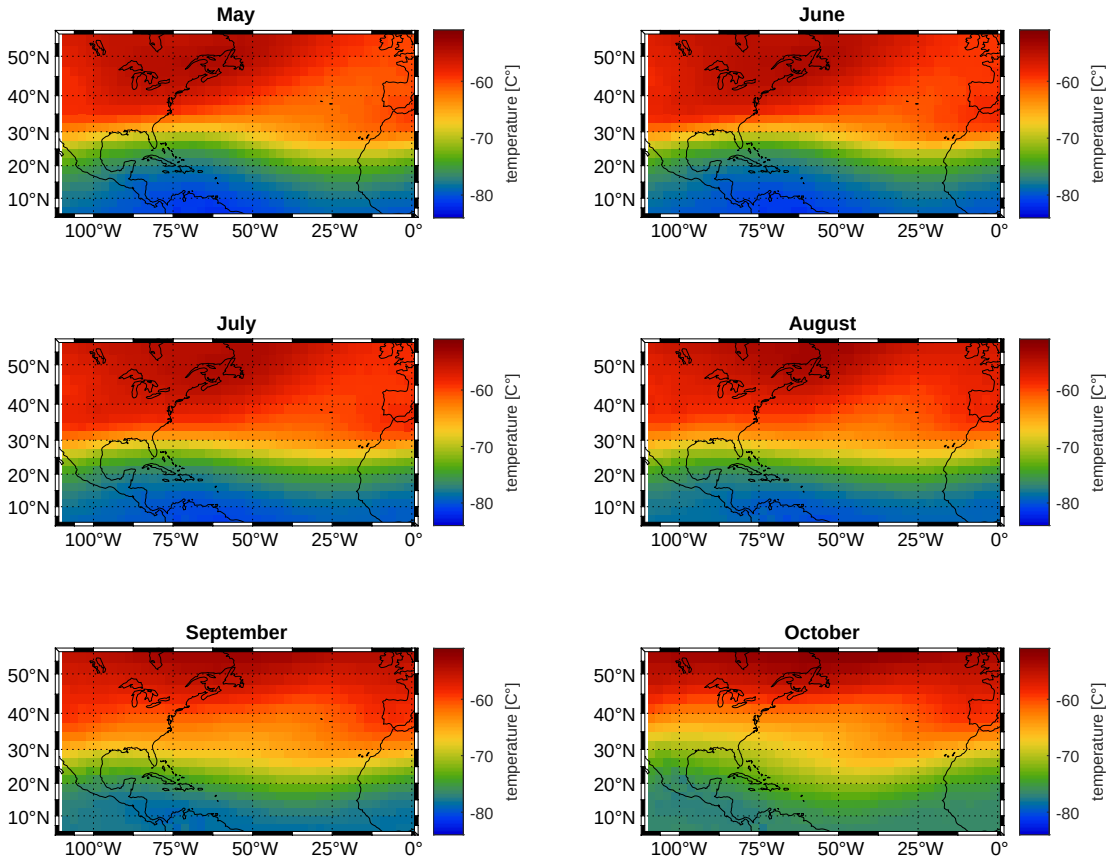


Figure 5.19: Monthly average tropopause temperature in Celsius over the years of available data provided in the NCEP/NCAR database. Mean variations of the temperature over the Atlantic basin, Gulf of Mexico and Caribbean Sea are in display for the months May to October.

The temperature at the top of the troposphere at the storm center is determined from the NCEP/NCAR Reanalysis dataset, which provides monthly means for the period 1948 until the present on a $2.5^\circ \times 2.5^\circ$ geographical grid in degrees Celsius. The dataset was

obtained from www.esrl.noaa.gov and monthly tropopause temperatures were interpolated and averaged throughout the years for each month. Figure 5.19 displays the tropopause temperature for the months of May to October.

A significant constraint of the relative intensity simulation approach is the limited size of available central pressure data in the HURDAT database. For example, the post-1970 tracks have central pressure measurements for most of their time history increments, but 1,200 of the HURDAT track records have no information on central pressure. To address this issue and populate some of these historical track time series, a relationship between central pressure and maximum gradient wind speeds, when available, is utilized Darling (1991).

The formula for deriving central pressure from maximum wind is given by:

$$p_c = 1013 - \left(\frac{V_g - 5.843 + 0.558\,N}{14.118} \right)^2 \tag{5.23}$$

where N is the latitude north of the equator, V_g is the maximum gradient wind given by:

$$V_g = V_{max} - 1.5\,S^{0.63} \tag{5.24}$$

where S is the estimated translation speed (in knots), and V_{max} is the maximum wind speed as recorded in the HURDAT database which has to exceed 35 kt.

After calculating the relative intensity index at each time increment —when required information is available either directly from the recorded minimum central pressure or indirectly through converting the maximum recorded wind— all data are parsed to an easy-to-process data set array. Based on this array, each row refers to one data point and includes the following information:

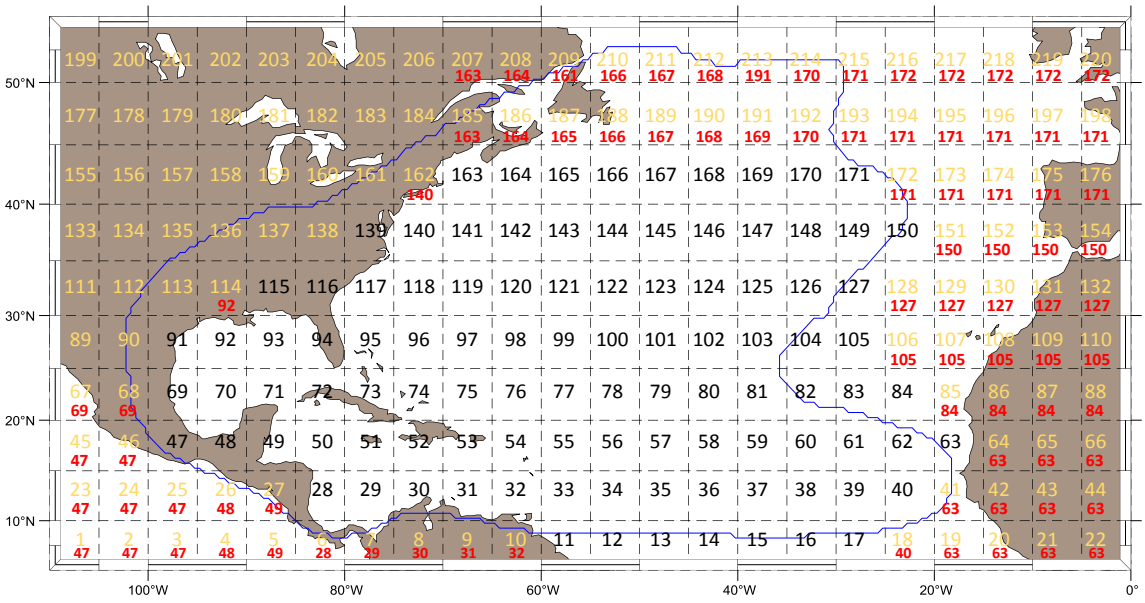


Figure 5.20: Atlantic basin mesh of the Westerly heading tropical cyclones. The cells with inadequate number of data get the relative intensity model coefficients from their neighbor cell numbered in red.

lat_i	lon_i	RI_{i+1}	RI_i	RI_{i-1}	RI_{i-2}	SST_{i+1}	SST_i	θ_i

Each data point is distributed to a geographic cell based on its geographic coordinates and further distributed to the Easterly or Westerly group based on the heading angle θ_i. For the cells with at least 7 data points, a regression analysis is performed and the coefficients described in Equation 5.13 are obtained. For the geographic cells which do not have at least 7 data points, the cells' coefficients are matched with those coefficients obtained from the closest cell with a minimum of 7 data points. Figure 5.20 shows the discretization of the Atlantic basin to $5^\circ \times 5^\circ$ grid cells and their reference number and depicts the results for the Westerly heading tropical cyclones. The cells without an adequate number of data points include an index in red, which which indicates the cell from which they match the regression coefficient. Figure 5.21 shows similar results for the Easterly heading storms.

The regression coefficients in Equation 5.13 are obtained from a least-square fit between the observed and the modeled values of the response. After the regression coefficients are

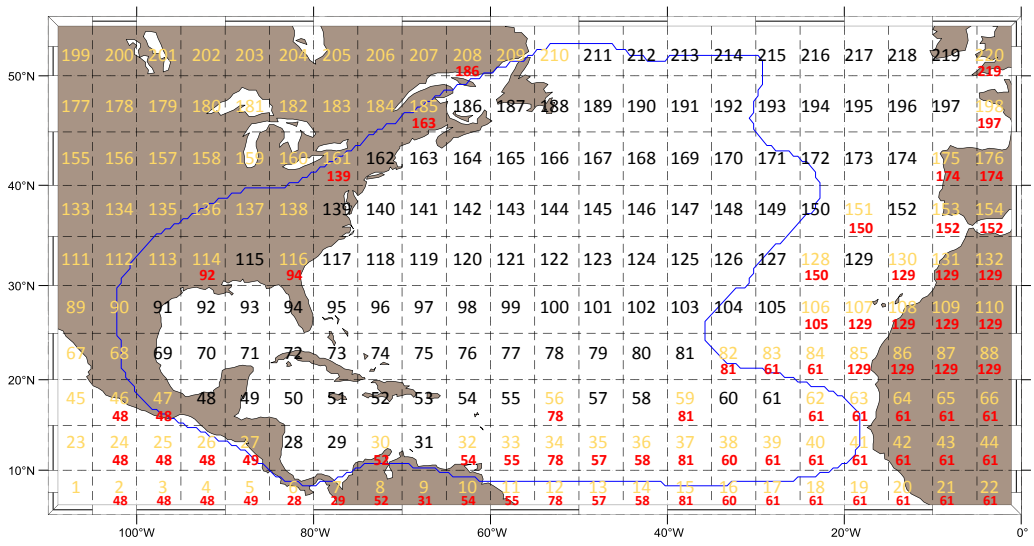


Figure 5.21: Atlantic basin mesh of the Easterly heading tropical cyclones. The cells with inadequate number of data get the relative intensity model coefficients from their neighbor cell numbered in red.

determined, the model response in each geographic cell is obtained and the difference (i.e., error) from the observations is calculated. The error is approximated separately with a normal and an unbounded Johnson distribution (Liu, 2014). The unbounded Johnson (UJ) distribution is given by:

$$z = \gamma + \delta \times \log\left\{\left(\frac{\varepsilon-\xi}{\lambda}\right) + \left[\left(\frac{\varepsilon-\xi}{\lambda}\right)^2 + 1\right]^{\frac{1}{2}}\right\}, \infty < \varepsilon < +\infty \qquad (5.25)$$

where z is the standard normal random variable, ε is the data population to be estimated, γ, δ are shape parameters, λ is the scale parameter and ξ is the location parameter. The UJ distribution fit to the data is made by matching 4 standardized normal quantiles, $[z_1 = -1.5, z-2 = -0.5, z_3 = 0.5, z_4 = 1.5]$. Having selected the evenly spaced quantiles, the standard normal cumulative probabilities at these values are computed first, and the quantiles of the elements in the data vector of the previously determined cumulative probabilities are computed next. Then, a system of four nonlinear equations is solved and the UJ distribution parameters $(\gamma, \delta, \xi, \lambda)$ are obtained. For all geographic cells the four-parametric

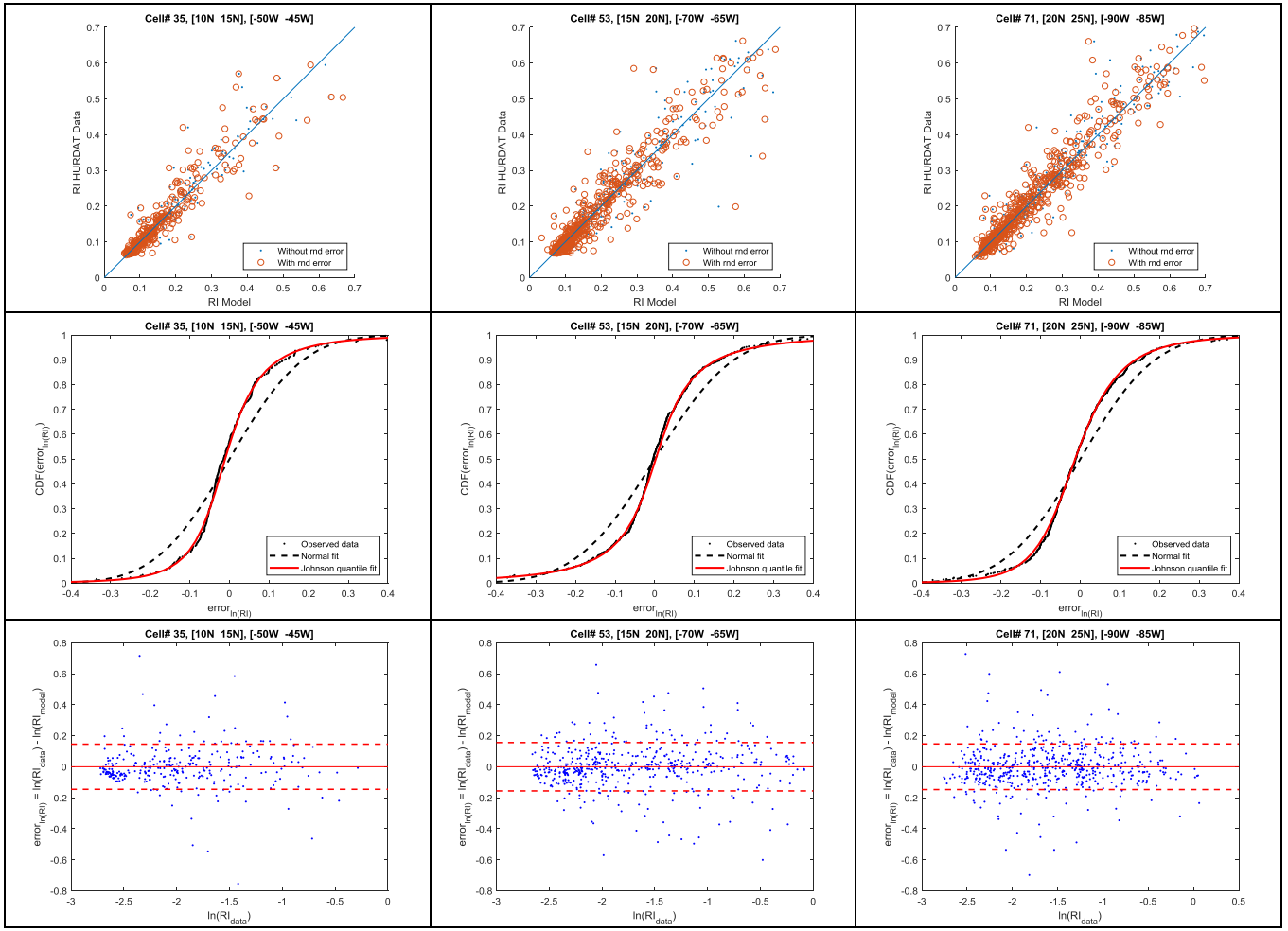


Figure 5.22: Errors of the relative intensity model for three geographic cells in the Atlantic basin, Caribbean Sea and Gulf of Mexico. From top to bottom: data response vs model response (top); empirical CDF of the of the errors with Normal distribution and Unbounded Johnson distribution fits (middle); scatter plot of logarithmic modeling errors (bottom).

Johnson distribution provides superior fit to the observational data.

Equation 5.13 demonstrates that information on the relative intensity index at the current and two previous positions is required to predict the relative intensity index of a anticipated geographic location. Post-1970 historical data was used to initiate the simulation of the track's relative intensity. The track with the closest genesis location to the simulated genesis location is selected form the historical record. The values of the relative intensity index at the fist three positions of the historic track are used as input in Equation 5.13. When there is more than one genesis location for a given historical record, one is selected at random. Lastly, the error ε_I in Equation 5.13 is simulated from the fitted Johnson S_U distribution and its values are constrained to lie within $\pm 1.5\,\sigma_{\varepsilon_I}$.

Equation (5.13) is used also for hurricanes which have already crossed the coastline and moved back to sea however, with the new grid-based intensity coefficients. To determine the relative intensity at the first location over water, the value of the central pressure obtained from the decay model (presented below) are used to estimate the relative intensity index which is used for all three priors required in Equation (5.13).

5.3.2 Assessment of relative intensity model

Figure 5.22 depicts errors of the relative intensity model for three geographic cells of the Westerly tropical cyclones. Similar results are also obtained for Easterly storms. The observed and modeled relative intensity fall around the first diagonal, indicating an unbiased model fit. The same conclusion can be drawn from the scatter plot of the relative intensity model errors which are distributed around zero. The cumulative distribution function of the errors implies that its distribution can be very well-approximated with an unbounded Johnson distribution.

5.4 Overland Filling Rate

A key component of the tropical cyclone simulation model is the modeling of the decay of the storm's intensity as the eye of the tropical cyclone crosses the coastline (i.e., makes landfall). Once a storm makes landfall it weakens and its central pressure rises. The rate of the storm's intensity loss is a function of the following variables, among others: geographic location at landfall, local topology and climatology, radius to maximum wind velocity, and translation velocity. Although these aspects are clearly understood, this wasn't the case 40 years ago when one of the first hurricane filling rate models was introduced by Schwerdt et al. (1979) who realized the importance of the location the storms makes landfall on its rate of intensity degradation. As the amount and the quality of full-scale data available increased substantially over the following years, many researchers developed empirical filing rate models as a function of multiple physical parameters. A collection of some of the most characteristic publications proposed filling rate models and their key components are presented in Table 5.2.

5.4.1 Filling rate model

This chapter utilizes the Vickery (2005) model which models the pressure decay after landfall as a function of the storm's intensity, size and translation speed at landfall, time after landfall and the landfalling region. The four landfall regions along with the geographic coordinates of their starting and ending locations along the coastline are depicted in Figure 5.23. The regions in order of Westerly to Easterly are the Gulf Coast, the Florida Peninsula, the min-Atlantic and the region of New England. The filling rate model is an exponential function yielding the difference of the central pressure and the far field pressure

Table 5.2: Filling rate models and their key characteristics

	Dependent Variable		Independent Variable							
	Central Pressure	Wind speed	Intensity @ landfall	Max. wind @ landfall	Regional inf.	Angle @ landfall	RWM	Storm speed	t after landfall	d. from landfall
Batts et al. (1980)	X	-	X	-	-	X	-	-	X	-
Georgiou (1986)	X	-	X	-	X	-	-	-	-	X
Vickery (1995)	X	-	X	-	X	-	-	-	X	-
Kaplan (1995)	-	X	-	X	-	-	-	-	X	-
vickery (2005)	X	-	X	-	X	-	X	X	X	-

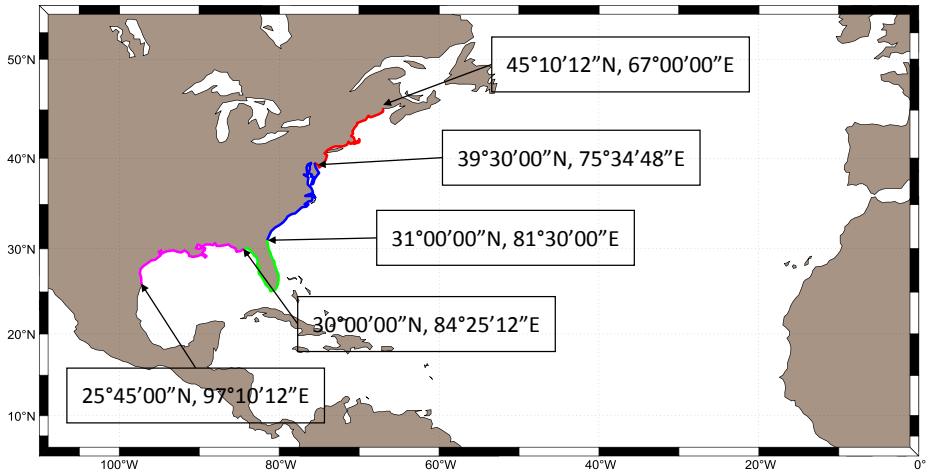


Figure 5.23: The four geographic regions described by the decay constant α (Equations 5.27, 5.28) and the regression parameters provided in Table 5.3. The coastline with magenta color refers to the Gulf Coast, the green color coastline refers to the Florida Peninsula, the blue coastline refers ro the Mid-Atlantic Coast and the red coastline refers to the New England Coast.

after landfall, ΔP_c in millibars [mb] and is given by:

$$\Delta P_c(t) = \Delta p_o \ \exp(-\alpha t) \tag{5.26}$$

where Δp_o is the difference between the central pressure of the storm and the far field pressure at the time the storm makes landfall, $\Delta p_o = P_c - P_o$, t is the time after landfall in unit of hours and α is the filling constant. In the case of the Gulf Coast, the Florida Peninsula and the mid-Atlantic Coast, the filling constant α is modeled as:

$$\alpha = \alpha_0 + \alpha_1 \left(\Delta p_0 \ c/RMW\right) + \varepsilon_\alpha \tag{5.27}$$

where c is the storm's translation speed at the time of landfall in [m/s], RMW is the radius to maximum wind in [km], ε_α is the modeling error resulting from the regression analysis, α_0 and α_1 are region-specific constants resulting from the linear regression analysis. In the

case of New England, the filling constant α is modeled as:

$$\alpha = \alpha_0 + \alpha_1 \Delta p_o + \varepsilon_\alpha \tag{5.28}$$

The values of the filling rate model constants and the modeling error parameter are given in Table 5.3 along with the sample size (N) used in the regression (Vickery, 2005). The far field pressure equals 1,013 mb. The storm dissipates when the central pressure of the eye reaches 1012 mb.

The central pressure at the time of landfall is determined and used as input in the filling rate models (Equations 5.26, 5.27 and 5.28). Landfall central pressure is equivalent to the last simulated oceanic central pressure. The time the storm spent over land until it reaches the first simulated overland position is considered proportional to the overland distance covered during the 6-hourly increment. This is because the storm's speed is considered constant between subsequent locations. The minimum allowable value of a sampled filling coefficient ε_α is set to 0.015, with the sampled error limited to lie within $\pm 3\,\sigma_{\varepsilon_\alpha}$.

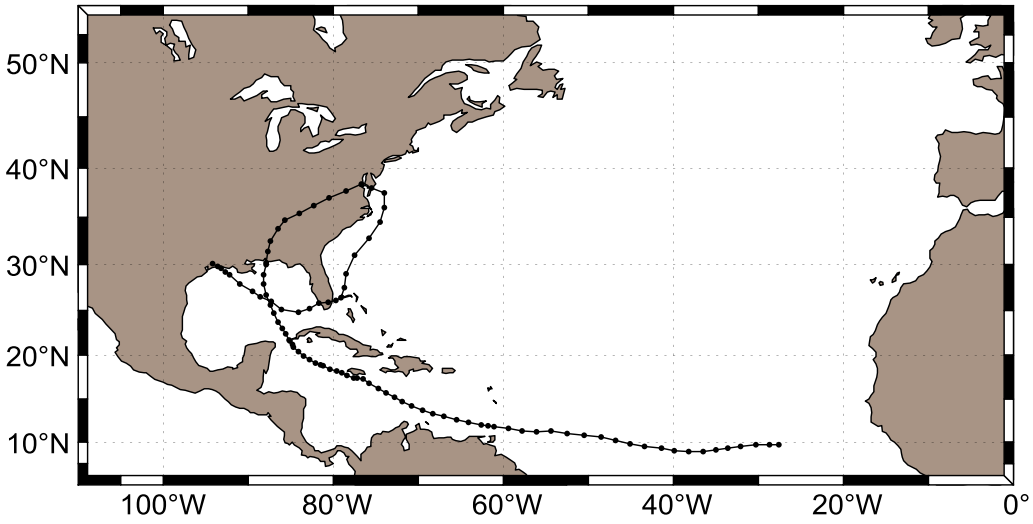


Figure 5.24: Track of hurricane Ivan occurred in 9/2/2004.

For the tracks which cross the coastline Southern to the Gulf Coast region (Figure 5.23), such as hurricane Harvey depicted in Figure 5.26, it is assumed that their filling constant

Table 5.3: Decay constant α, regression parameters (RMW in [km]; translation speed c, in [m s^{-1}]; and Δp_o, in [mb]; α_0 is the intercept and α_1 is the slope).

		$\alpha = \alpha_0 + \alpha_1 \Delta p_o$				$\alpha = \alpha_0 + \alpha_1 \Delta p_o c/RMW$			
Landfall region	N	α_1	α_0	r^2	σ_ε	α_1	α_0	r^2	σ_ε
Gulf Coast	26	-	-	-	-	0.00181	0.0414	0.5884	0.0169
Florida Peninsula	13	-	-	-	-	0.00167	0.0225	0.8378	0.0158
Mid-Atlantic Coast	13	-	-	-	-	0.00156	0.0370	0.4206	0.0161
New England Coast	6	0.00099	0.0034	0.5471	0.0114	-	-	-	-

follows the model described by Equation 5.27.

Figure 5.24, and 5.26 depict the tracks of 2004 Hurricane Ivan and 2017 Hurricane Harvey, respectively. Both of these hurricanes made landfall at the Gulf Coast. Figures 5.25 and 5.27 show 200 simulated samples of the central pressure of hurricanes Ivan and Harvey, respectively.

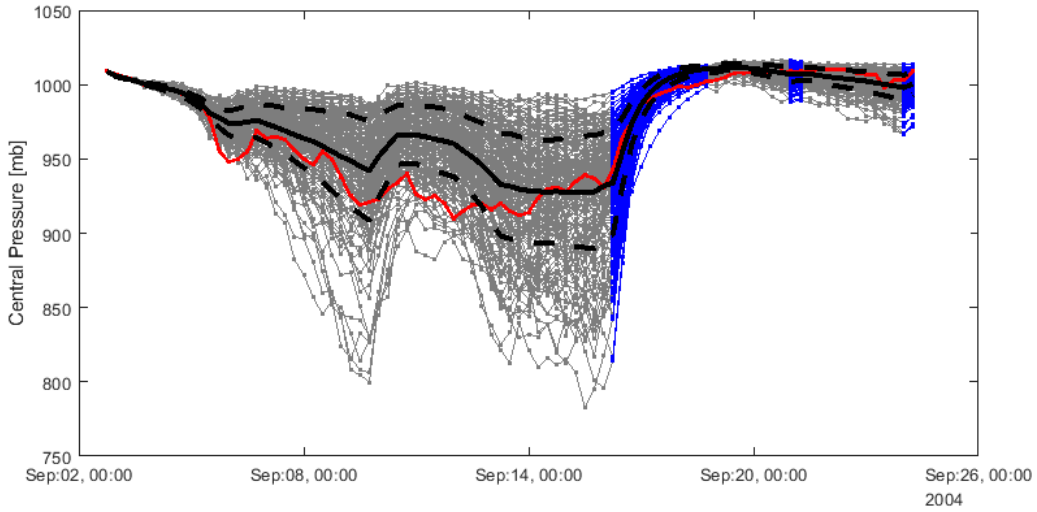


Figure 5.25: Simulated central pressures of hurricane Ivan formed in September 2, 2004. The red solid line shows the observed central pressure, in solid black line is the mean simulated value and the dashed black lines depict $\pm 1\,\sigma$. The gray lines are showing the central pressure values over water for different samples and the blue segments refer to the overland values.

5.5 Radius to maximum wind model

Radius to maximum wind (RMW) is the most important determinant of a tropical cyclone's size. The first empirical models of the RMW were developed by Vickery and Twisdale (1995). Vickery et al. (2000a) performed the same analysis using 8 more storms and updated the regression equations for prediction of RMW. As more accurate and better resolution data became available for the storm radius by the National Weather Services (NWS) and by the Hurricane Research Division's H*Wind snapshots of storm's wind fields, new statistical models were developed. Vickery and Wadhera (2008) used the updated database

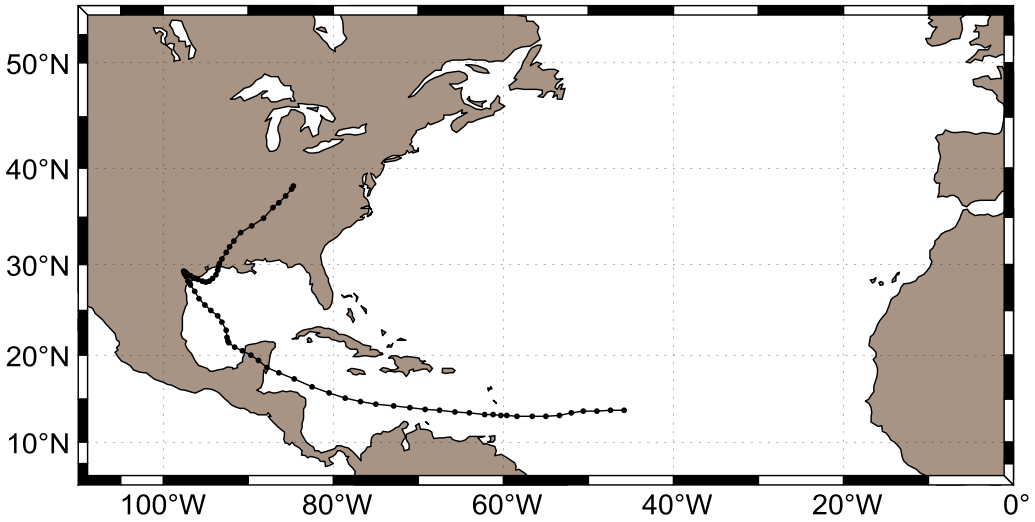


Figure 5.26: Track of hurricane Harvey occurred in 8/16/2017.

until 2005 and developed new statistical models relating RMW to latitude and central pressure, and compared them with the regression models derived from U.S. land falling storms alone.

Three RMW models were developed for hurricanes in the Gulf of Mexico, in the Atlantic ocean, and in both oceans combined (Vickery and Wadhera, 2008). The RMW models used in this chapter are following the guidelines developed by Vickery et al. (2009b) and part of the models developed in Vickery and Wadhera (2008). Specifically, the all-storms RMW model developed in Vickery and Wadhera (2008) is applied to the Atlantic ocean hurricanes and the Gulf of Mexico RMW model is applied to the hurricanes in the Gulf of Mexico. The RMW model applied ot the Gulf of Mexico storms is given by:

$$\ln(\mathrm{RMW}_{\mathrm{Gulf}}) = 3.859 - 7.700 \cdot 10^{-5} \Delta p^2 + \varepsilon_{\mathrm{Gulf}} \tag{5.29}$$

and the error $\varepsilon_{\mathrm{Gulf}}$ is approximated with Normal distribution, $N(0,1) \cdot \sigma_{\mathrm{lnRMW}}$, where

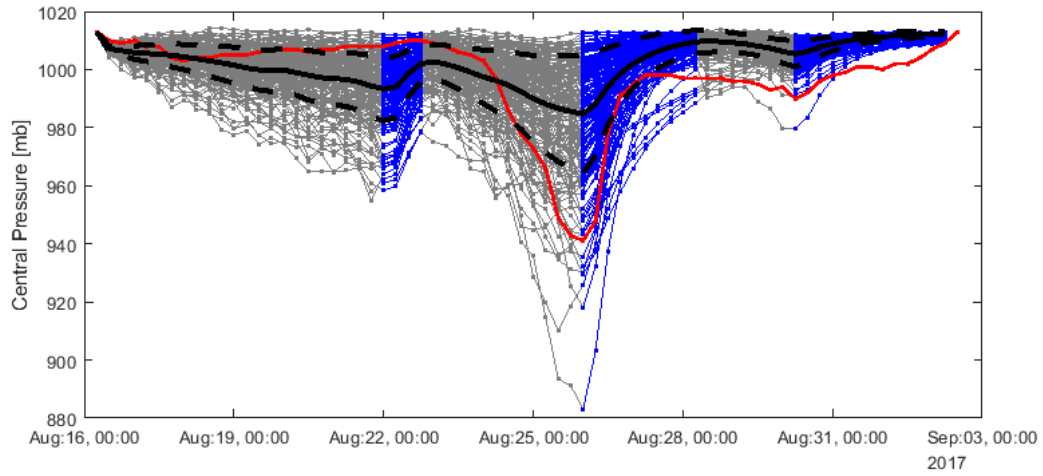


Figure 5.27: Simulated central pressures of hurricane Harvey formed in August 17, 2004. The solid red line shows the observed central pressure. The solid black line is the mean simulated value. The dashed black lines depict $\pm 1\,\sigma$. The gray lines are showing the central pressure values over water for different samples and the blue segments refer to the overland values.

σ_{lnRMW} is given by:

$$\begin{aligned} \sigma_{\text{lnRMW}} &= 0.396; && \Delta p \leq 100hPa, \\ \sigma_{\text{lnRMW}} &= 1.424 - 0.01029\Delta p; && 100hPa \leq \Delta p \leq 120hPa \\ \sigma_{\text{lnRMW}} &= 0.19; && \Delta p > 120hPa \end{aligned} \tag{5.30}$$

The RMW model for Atlantic hurricanes is given by:

$$\ln(\text{RMW}_{\text{Atlantic}}) = 3.015 - 6.291 \cdot 10^{-5}\Delta p^2 + 0.0337\Psi + \varepsilon_{\text{Atlantic}} \tag{5.31}$$

and the error $\varepsilon_{\text{Atlantic}}$ is approximated with Normal distribution, $N(0,1) \cdot \sigma_{\text{lnRMW}}$, where σ_{lnRMW} is given by:

$$\begin{aligned} \sigma_{\text{lnRMW}} &= 0.448; && \Delta p \leq 87hPa, \\ \sigma_{\text{lnRMW}} &= 1.137 - 0.00792\Delta p; && 87hPa \leq \Delta p \leq 120hPa \\ \sigma_{\text{lnRMW}} &= 0.19; && \Delta p > 120hPa \end{aligned} \tag{5.32}$$

where Δp is the central pressure difference in [mb] and Ψ is the latitude in [deg].

To yield one statistical model of the RMW applied to the simulated storm, the two models described by Equations 5.29 and 5.31 are combined as follows:

$$\begin{aligned} \text{RMW} &= \alpha_1 \text{RMW}_{\text{Atlantic}} + (1-\alpha_1)\text{RMW}_{\text{Gulf}} \\ \alpha_1 &= \frac{\Sigma \Delta p_{\text{Atlantic}}}{\Sigma[\Delta p_{\text{Atlantic}} + \Delta p_{\text{Gulf}}]} \end{aligned} \tag{5.33}$$

where the summation is performed over all six-hour time steps from storm genesis to the current time.

For the simulation of both errors $\varepsilon_{\text{Atlantic}}$ and $\varepsilon_{\text{Gulf}}$ a common random number is drawn from a $N(0,1)$ distribution and then scaled with the corresponding error's standard deviation. This is performed only once in the beginning of each simulated storm and the error value is used throughout the simulation.

As will be seen later, the RMW has no effect on the magnitude of the maximum wind speeds; however, it has a significant influence on the area affected by a tropical cyclone. For a single-site wind risk study, the modeling of the RMW impacts the likelihood of the site experiencing strong winds in cases of near misses. Accurate modeling of the RMW is critical to storm surge, wave and wind field modeling, as well as to the estimation of probable maximum losses for insurance modeling purposes. RMW magnitudes are inversely correlated with central pressure differences, such that RMW decreases as storm intensity increases, and vice-versa. The RMW also increases with increasing latitudes.

5.6 Holland B parameter

In the hurricane hazard modeling, the inclusion of the Holland B parameter plays an important role in the hurricane risk studies. In 1980, Holland proposed a representation of the gra-

dient hurricane wind field by introducing an additional parameter to define the maximum wind in a hurricane. in an effort to differing central and far field pressures, Holland initially modeled hurricane radial surface pressure profiles using parameter $(p - p_0)/(p_n - p_0)$, where p is the pressure at radius r, p_0 the central pressure and p_n the far field pressure. He approximated the radial profiles of this parameter with a rectangular hyperbola function, which in a rearranged form is given by:

$$p(r) = p_0 + (p_n - p_0)\exp\left(-\frac{A}{r^B}\right) \tag{5.34}$$

It has been shown that RMW $= A^{1/\mathrm{B}}$ and thus Equation 5.34 becomes:

$$p(r) = p_0 + (p_n - p_0)\left[\exp\left(-\frac{RMW}{r}\right)\right]^B \tag{5.35}$$

Parameter B in Equation 5.35 is the Holland B parameter, which defines the shape of the radial profile. Equation 5.35 and Holland B parameter have since their introduction been used by many researchers as part of their wind field models (Georgiou, 1986; Thompson and Cardone, 1996; Vickery et al., 2000b). Vickery et al. (2000a) used Holland B as random variable for first time in his hurricane simulation methodology where he estimated B using reconnaissance aircraft data and modeled it as a function of RMW and central pressure difference (Δp).

5.6.1 Holland B parameter: Oceanic model

The latest statistical model for Holland B, which is utilized herein, was developed by Vickery and Wadhera (2008) introducing the latitude term in their model. The authors compared B values estimated from reconnaissance and H*Wind data with those estimated using a

wind field model and surface level wind speed measurements. Parameter B is given by:

$$B = 1.7642 - 1.2098\sqrt{A} + \varepsilon_B \tag{5.36}$$

where A is a nondimensional variable and ε_B is the model error. Variable A is defined by:

$$A = \frac{RMW f_c}{\sqrt{2R_d(T_s - 273)\ln\left(1 + \frac{\Delta p}{p_c e}\right)}} \tag{5.37}$$

where RMW is the radius to maximum wind in [m], f_c is the Coriolis force, defined as $2\Omega\sin\psi$, where Ω is the rotational rate of the earth given as $7.2921 \cdot 10^{-5}$ in [rad/sec], ψ is the latitude in [degrees], R_d is the gas constant for dry air given as 287 [N m / (kg K)], T_s is the mean sea surface temperature in [kelvin], p_0 is the pressure at the center of the storm in [mb], Δp is the difference between p_0 and the far field pressure taken here as 1,013 [mb] and e is the base of the natural logarithms. Parameter ε_B is the model error approximated with normal distribution, $N(0, \sigma_B^2)$ and $\sigma_B = 0.226$. This error is sampled only once prior to the start of each simulated storm and is used throughout the simulation as a shift from the regression model. From Equations 5.36 and 5.37 the following observations are made:

(i) as RMW decreases B increases

(ii) as latitude increases $\sin\psi$ increases and thus B decreases

(iii) as T_s increases B decreases

(iv) as the ratio $\frac{\Delta p}{p_c}$ increases B decreases

5.6.2 Holland B parameter: Land model

The Holland B model described in Equations 5.36 and 5.37 has been derived for the simulation of tropical cyclones over ocean. Vickery et al. (2009b) proposed a decay model in

which parameter B after landfall is given by:

$$B(t) = B_0 \exp(\alpha\, t) \tag{5.38}$$

where parameter α is given by:

$$a = 0.0291 - 0.0429\, B_0; \quad a \leq -0.005 \tag{5.39}$$

where B_0 is the value of B at landfall and t is the time the storm spent over land in [h].

5.7 Gradient wind field

Wind field models differ in the input they require, the output provided, implementation difficulty, and computational resources needed to accurately model a storm's wind field along its path. In this chapter, the choice of the wind field model is made based on the idea to develop joint probability densities of wind intensity and directionality at many geographic locations. Therefore, an effective yet computationally efficient algorithm should be selected and thus the gradient wind field model developed by Georgiou (1986) is adopted. The gradient wind equation considered in Georgiou's model describes the balanced atmospheric flow at the 700 [mb] level (or 3,000m height), which is given by:

$$\begin{aligned} \frac{1}{\rho}\frac{\partial p}{\partial r} &= \frac{V^2}{r} + fV \\ V_{rad} &<< V \end{aligned} \tag{5.40}$$

where, V, V_{rad} are the tangential and radial velocity components, ρ is the air density, f is the Coriolis parameter, p is the pressure and r is the distance from the storm center. Equation (5.40) assumes a steady state flow and therefore the mean wind speed of the air parcels are computed as an outcome. Georgiou (1986) considered also the effect of the

storm's translation speed (VT) using the Blaton's formula which is given by:

$$\tilde{r} = \frac{r}{1 - \frac{VT}{V}\sin(\alpha)} \tag{5.41}$$

where VT is the storm's translation speed and α is the angle, clockwise positive, from the

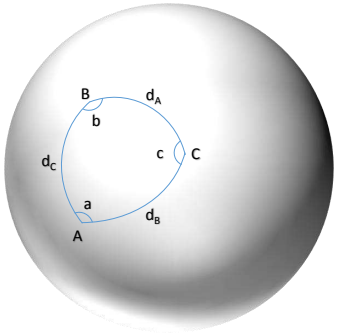


Figure 5.28: Schematic of three locations over a sphere with their great cycle distances.

storm translation direction to the target location. Using Equation 5.35 in Equation 5.40 and replacing the radial distance r in Equation 5.40 with Equation 5.41, the gradient balance wind field is thus given by:

$$V(r,\alpha) = \frac{1}{2}(VT\sin(\alpha) - fr) + \sqrt{\frac{1}{4}(VT\sin(\alpha) - fr)^2 + B\frac{\Delta p}{\rho}\left(\frac{RMW}{r}\right)^B \cdot \exp\left[-\left(\frac{RMW}{r}\right)^B\right]} \tag{5.42}$$

Angle α can be computed using spherical trigonometry and is considered clockwise positive with range $[0°, 360°]$. Given three geographic locations A, B and C and their in between great circle distances d_A, d_B and d_C as depicted in Figure 5.28, the angle a can be computed by:

$$a = \cos^{-1}\left[\frac{\cos(d_A) - \cos(d_B)\cdot\cos(d_C)}{\sin(d_B)\cdot\sin(d_C)}\right] \tag{5.43}$$

Figure 5.29 shows a track making landfall at the Gulf Coast and the simulated gradient winds at 3,000m before the eye crosses the coastline.

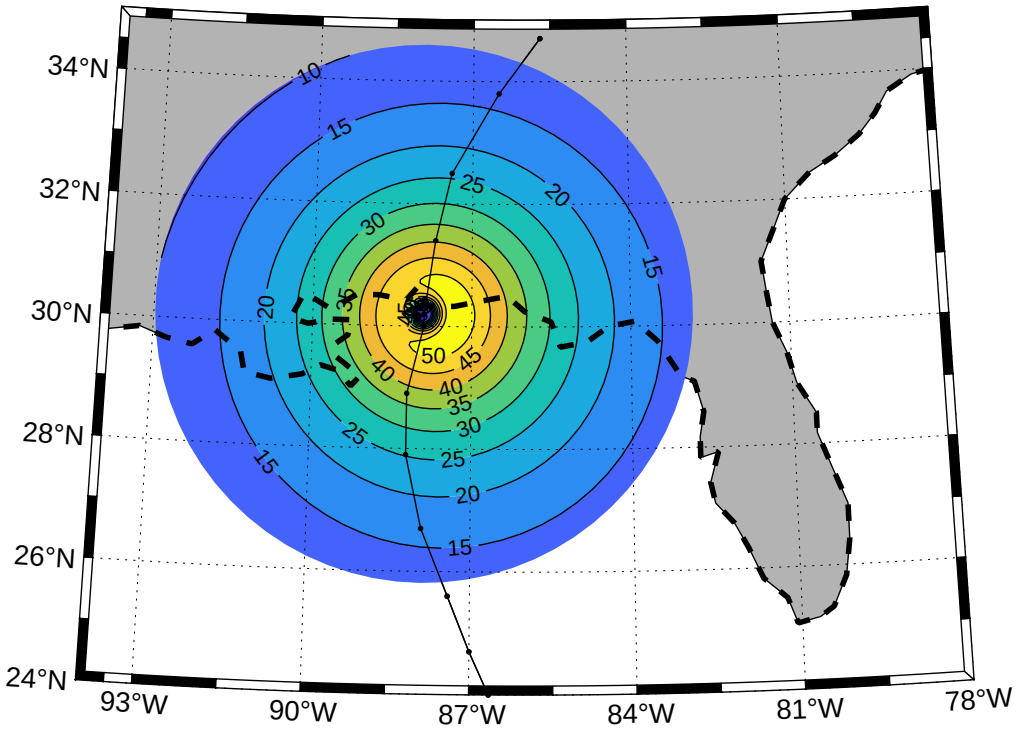


Figure 5.29: Simulated gradient 1-min sustained wind speed at 3,000 m height. The track and central pressure information is adopted from Hurricane Ivan (2004) at 650 UTC. The gradient wind has been simulated suing a sample of RMW = 61.69 km and a sample of Holland B equal to 1.256. The path is depicted with solid black line and the dashed black line shows the coastline. The wind intensity is in [m/s].

The asymmetric wind field given by Equation 5.42 which can be seen as a translation vortex, is a function of four parameters: RMW, Holland B parameter, central pressure difference from far field pressure Δp and storm's translation speed VT.

5.8 Boundary layer model

The boundary layer model (BLM) describes the profile of the mean wind speed over the height of a hurricane. When the mean wind speed at the gradient height is estimated, the atmospheric boundary layer model provides the wind speed at any height down to the surface of 10 m above the ground or water. These semi-empirical models are providing more information and higher accuracy at the surface wind speed than the more commonly

used speed reduction factors. The models developed by Vickery et al. (2009a) are used herein for the estimation of the over water and land wind speeds. This hurricane boundary layer model reproduces the shape of the hurricane boundary layer for heights between 1,000 and 10 m. The model, besides the height information, account also for the transition and shift of the roughness regime from open water to overland and if given by:

$$U(z) = \frac{u_*}{k}\left[\ln\left(\frac{z}{z_0}\right) - \alpha\left(\frac{z}{H^*}\right)^n\right] \tag{5.44}$$

where $U(z)$ is the wind speed at height z above the surface of the land or water, k is the Kármán coefficient having a value of 0.4, u^* is the friction velocity, z_0 is the surface roughness in [m], parameters α and n are having values equal to 0.4 and 2, respectively and H^* is the boundary layer height parameter.

5.8.1 Oceanic Boundary layer model

For the over water (marine) boundary layer, the boundary layer height parameter H^* is given by:

$$H^* = 343.7 + 0.260/I + \varepsilon_H \tag{5.45}$$

where ε_H is Normally distributed with zero mean and $\sigma_{\varepsilon_H} = 99$ m (constraint within $2\sigma_{\varepsilon_H}$) and I is the inertial stability. The inertial stability (Kepert, 2001; Vickery et al., 2009a) is given in by:

$$I = \sqrt{\left(f + \frac{2V}{r}\right)\left(f + \frac{V}{r} + \frac{\partial V}{\partial r}\right)} \tag{5.46}$$

where V is the gradient wind speed at 1,000 m in [m/s], r is the radial distance from the storm's center in [m] and f is the Coriolis parameter.

To estimate the wind speed at one site, 10 m over water, we first estimate the friction velocity by sampling the boundary layer height parameter H^* using Equation 5.45 and

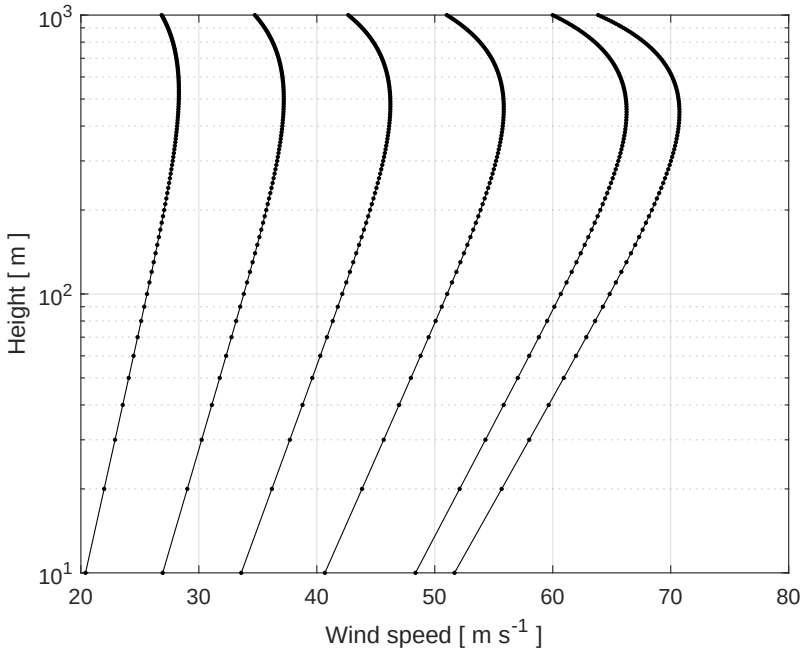


Figure 5.30: Marine boundary layer profiles for RMW = 20 [km], and velocities at that distance take the values [26.85, 34.75, 42.67, 51.02, 60, 63.58]. The surface roughness is taken as $z_0 = 0.0013$ and the latitude is assumed 32°N.

substitute in Equation 5.44 along with the estimated gradient wind speed at 1,000m. The friction velocity u_* is the only unknown and can be readily assessed. The wind speed at 10 m is assessed by replacing $z = 10$, the estimated u_* and sampled H^* in Equation 5.44. The water roughness z_0 is considered equal to 0.0013 m. Figure 5.30 shows boundary layer profiles for different wind speeds.

5.8.2 Land boundary layer model

For the sea-land transition, the wind speed at the top of the boundary layer, considered here 1,000, is assumed to remain unchanged as the flow moves over the new roughness regime. To estimate the wind speed at a sit, 10 m over open terrain, a similar approach as previously explained for the over water case is followed. The boundary layer height parameter H^* is determined by first computing the jet height or boundary layer height H

according to Kaplan and DeMaria (2001):

$$H = \sqrt{\frac{2K}{I}\left[\pi + \tan^{-1}\left(-1 - \frac{2}{\chi}\right)\right]}$$
$$\chi = C_d V \sqrt{\frac{2}{KI}} \tag{5.47}$$

where C_d is the drag coefficient, K is the turbulent diffusivity of momentum taken as constant and equal to $k = 75$ [m^2 / s^{-1}] (Vickery et al., 2009a).

Parameter H^* which is then given by $H^* = H/1.12$ (Vickery et al., 2009a), is then substituted into Equation 5.44 along with the estimated gradient wind speed and Equation 5.44 is solved for u_*. Parameter u_* is back substituted with H^* into Equation 5.44 and for $z = 10$ m the wind speed at 10 m height over land is estimated. The surface roughness of an open terrain is taken equal to $z_0 = 0.03$ m which corresponds to a $C_d = 0.00475$ (Powell, 2008).

Figure 5.31 displays surface 1-min sustained wind speed at 10 m height for a track prior to landfall.

5.9 Wind directionality

Hurricane wind directionality and the inflow angle are outcomes of the storm's characteristics and the relative location of a geographic position of interest with respect to the storms center and direction. The inflow angle at a geographic position refers to the angle between the true wind direction and a tangent to a circle with center at the storms center. It has been identified that the inflow angle depends upon many characteristics of the hurricane such as the surface roughness, the storm quadrant, storm's transnational velocity, latitude, and storm's intensity and radius to maximum wind (Dunn and Miller, 1964; Simpson and Riehl, 1981). However, not many studies have been developed on inflow angles of tropical cyclones –especially over land (Zhang, 2019)– mainly due to the lack of observations with

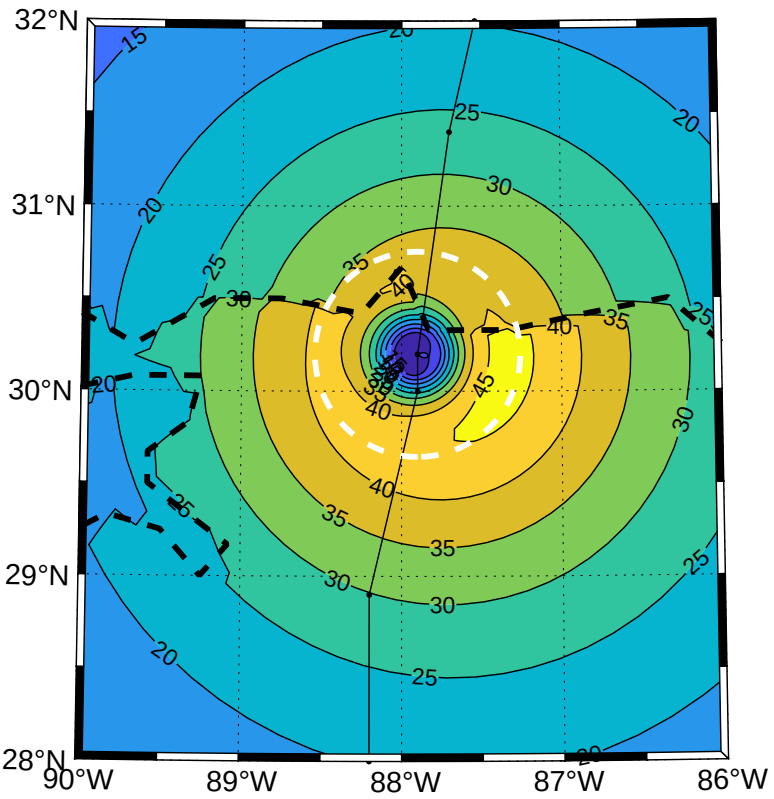


Figure 5.31: Simulated surface 1-min sustained wind speed at 10 m height. The track and central pressure information is adopted from Hurricane Ivan (2004) at 650 UTC. The gradient wind has been simulated using a RMW of 61.69 km and Holland B equal to 1.256. The path is depicted with solid black line, the dash black line shows the coastline and the white dashed line represents the RMW. The wind intensity is in [m/s] and the higher wind speed reduction over land is clearly depicted along the coastline. It should be noted that no fetch distance was considered in this simulation.

good coverage and accuracy.

Historical wind direction observation had been very sparse even compared to the surface wind speed data (Zhang and Uhlhorn, 2012). The reported observations in the literature were initially only enough to provide mean estimates of the inflow angles. Graham and Nunn (1959) reported inflow angles changing as a function of the distance from the storm. The authors suggested a mean inflow angle of 20° for locations within the radius to maximum wind RMW, $20° - 25°$ for locations within $RMW - 1.2RMW$ and 25° for locations $1.2RMW$ and beyond. Dunn and Miller (1964) reported angles of average inflow to

be 23° over water and 38° over land. Whereas, other literature reported mean values over water between $15° - 20°$ for moderate hurricanes and $20° - 40°$ for severe hurricanes plus an additional $10° - 30°$ over land (Simpson and Riehl, 1981). The most cited studies on observations on inflow angles are publications by Powell M.D. and his co-workers. Powell (1982) reported inflow angles to range between $+12°$ and $-55°$ over open water for hurricane Frederic, with greater inflow in the right-rear quadrant and weaker inflow in the left-front quadrant.

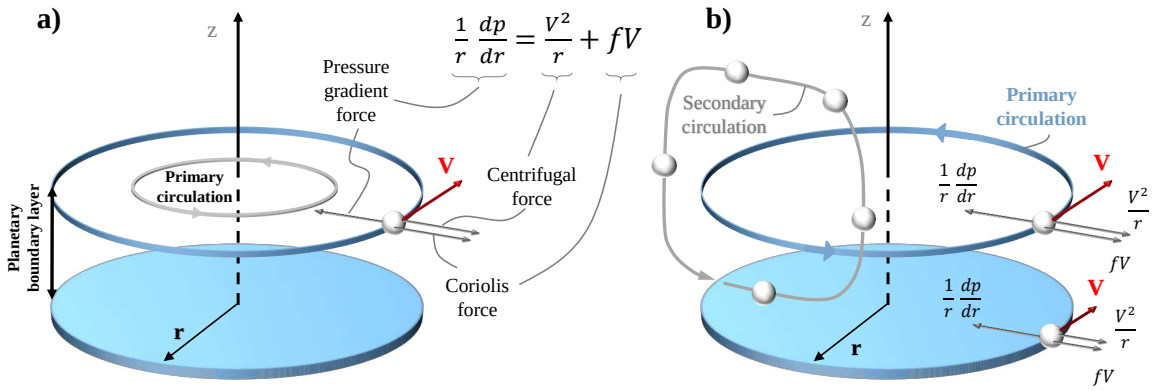


Figure 5.32: Schematic structure of a steady state flow in a mature storm, where in a) the primary circulation and the gradient wind balance is seen and in b) and the secondary circulation with the frictionally-induced convergence.

The primary reason of the inflow angle is the effect of frictional force occurring in the boundary layer close to the ground. For better understanding, the flow structure in a mature hurricane shall be examined. The spiralling flow observed in a tropical cyclone is a result of two transverse circulations, the primary circulation depicted in Figure 5.32 a) which is horizontal and the secondary circulation depicted in Figure 5.32 b) which is vertical. The forces act on an air parcel, which is quite distant from the ground, are the pressure gradient force, which is due to the pressure drop directing the air parcel inwards towards the storm center, and, the centrifugal and Coriolis forces which act to move the parcel outwards and balance the pressure gradient force. This balance is known as the gradient wind balance and is depicted in Figure 5.32 a). For air parcels located closer to the ground, the surface friction starts appearing, with increasing magnitude as getting closer to the ground, and

acts additionally to the rest of the forces.

The surface friction forces turn out to have a significant influence on the dynamics of the tropical cyclone. More specifically, the frictional force reduces the tangential wind speed near the surface and as a result it reduces the centrifugal and Coriolis forces as depicted schematically in Fiugre 5.32. However, the effect of friction is limited on the pressure field in the boundary layer and the redial pressure gradient force is approximately the same as that immediately above the layer. Therefore, the pressure gradient force acting on a parcel of air is almost the same throughout the whole depth of the layer. The consequence is that by getting closer to the surface, the pressure gradient force is higher than the superposition of the centrifugal and Coriolis forces and the parcel of air is leaning inward towards the center of the storm. The resulting inflow angle is stronger close to the ground and a height of about 50 to 100 m (Vogl, 2009).

Traditionally, when wind directionality was studied parametrically, the inflow angles were considered to take a fixed value contingent upon the location being over land or water and its distance from the storm's center being within the radius to maximum wind or exceeding it (Sanchez-Sesma et al., 1988; Boose et al., 1994). The latest observation-based parametric model for sea surface inflow angle was proposed by Zhang and Uhlhorn (2012). The author proposed a model for the spatial distribution of the near-surface inflow angle which is composed of two parts, the axisymmetric part of mean values, and the asymmetric part. The model is accounting for the storm motion speed, intensity, size and the relative position of the location of interest with respect to the storm's center and direction.

5.9.1 Over water paramteric inflow angle model

The parametric inflow angle proposed by Zhang and Uhlhorn (2012) is utilized in this chapter for the over water locations. The parametric model of the storm-relative inflow

angle α_{SR} is given by:

$$\alpha_{SR}(r^*, \theta, V_{max}, V_s) = A_{\alpha 0}(r^*, V_{max}) - A_{\alpha 1}(r^*, V_s) \cdot \cos\left[\theta - P_{\alpha_1}(r^*, V_s)\right] + \varepsilon_{\alpha_{SR}} \quad (5.48)$$

where r^* is the normalized radial distance given by $r^* = r/RMW$, θ is the storm-relative azimuthal direction (i.e., clockwise positive from the bearing direction), V_{max} is the storm's maximum wind intensity in [m/s], V_s is the storm's motion speed in [m/s]. The axisymmetric inflow angle $A_{\alpha 0}$ is given by:

$$A_{\alpha 0} = a_{A0} r^* + b_{A0} V_{max} + c_{A0} \quad (5.49)$$

where coefficients a_{A0}, b_{A0} and c_{A0} are given in Table 5.4. The asymmetric model is a harmonic function consisting of a mean plus wavenumber-1 component described by the normalized amplitude $A_{\alpha_1}/A_{\alpha_0}$ and phase P_{α_1}. The normalized amplitude is given by:

$$\frac{A_{\alpha_1}}{A_{\alpha_0}} = a_{A1}\, r* + b_{A1}\, V_s + c_{A1} \quad (5.50)$$

where coefficients a_{A1}, b_{A1} and c_{A1} are given in Table 5.4. The phase P_{α_1} is given by:

$$P_{\alpha 1} = a_{P1}\, r^* + b_{P1}\, V_s + c_{P1} \quad (5.51)$$

where coefficients a_{P1}, b_{P1} and c_{P1} are given in Table 5.4. The wind directionality α at a site of interest is determined by:

$$\alpha = \alpha_t - 90^\circ + \alpha_{SR} \quad (5.52)$$

where α_t is the angle between a reference vector pointing North to a vector pointing towards the site of interest starting from the storm's center, and, takes values in the closed interval [-π, π]. Parameters α_{SR} and α take values in the closed interval $[-\pi/2, \pi/2]$ and [-π,

Table 5.4: Mean μ and standard deviation σ values of the coefficients for the parametric inflow angle model.

Eq.	Variables		a	b	c
(5.49)	$A_{\alpha_0}(^\circ)$	μ	-0.90	-0.09	-14.33
		σ	0.29	0.07	4.22
(5.50)	$A_{\alpha_1}/A_{\alpha_0}$	μ	0.04	0.05	0.14
		σ	0.04	0.06	0.32
(5.51)	$P_{\alpha_1}(^\circ)$	μ	6.88	-9.60	85.31
		σ	5.80	9.42	56.86

π], respectively, and are depicted in Figure 5.33. The model yields mean inflow angle of

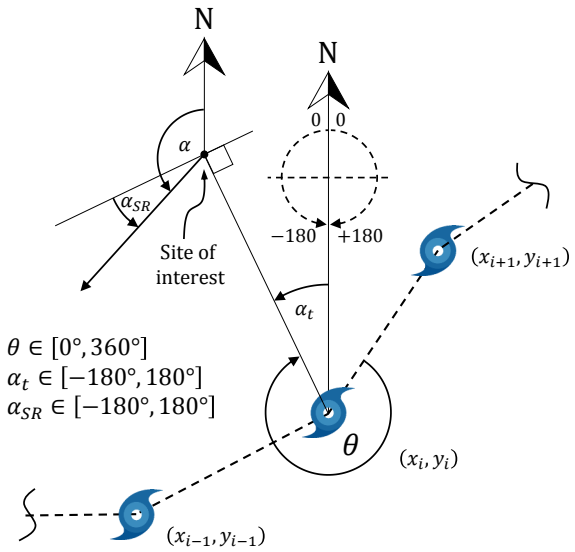


Figure 5.33: Schematic of the hurricane track and angle definition for site of interest.

$22.6^\circ \pm 2.2^\circ$ with 95% confidence (Zhang and Uhlhorn, 2012).

5.9.2 Over land paramteric inflow angle model

As already mentioned, the observational wind directionality records are too sparse over land even compered to the adequate wind directionality data over the ocean. Therefore, a parametric model does not exist to yield the distribution of the overland inflow angle and traditionally the wind direction estimation is then arrived at by applying a constant inflow angle to the counterclockwise tangential direction (Dunn and Miller, 1964; Boose et al.,

1994)

The most significant source of data used for the investigation of the near-surface inflow angle over a broad range of tropical cyclones is the dropwindsonde devices. These devices provide data with good accuracy which has not changed over the years and thus reliable studies can be performed (Powell et al., 2009; Zhang and Uhlhorn, 2012). However, dropwindsonde devices are used only over ocean. Other reasons for the limited amount of hurricane wind directionality data is that the traditional methods used to measure hurricane wind intensity are not providing reliable information on wind directionality. For example, the Nadir-viewing passive microwave radiometers (e.g., SFMR) are insensitive to wind direction (Zhang and Uhlhorn, 2012), active microwave sensors such as the NASA Quick Scatterometers (QuikSCAT) are affected by the rain contamination which complicates the interpretation of QuikSCAT wind direction retrievals (Brennan et al., 2009), and the synthetic aperture radar (SAR) provide single views and thus difficulties in assessing the wind direction (Shen et al., 2009).

Additionally, the surface weather stations are coarsely distributed and many do not survive the storm if their location is too close to the track of the tropical cyclone (Boose et al., 1994; Powell et al., 1996). Moreover, reconnaissance aircraft wind directionality data are limited over land because as Powell (1982) mentioned "the Air Force and NOAA reconnaissance aircraft are unable to penetrate the storm or fly over land because of the danger of added mechanical and convective turbulence and the possibility of tornadoes." In cases where the hurricanes approaches the land, the reconnaissance aircraft continue to provide information by flying parallel to the coast. The importance of retrieving hurricane wind directionality over land has increased past years and researchers are investigating on new techniques for extracting more accurate estimates from methods already used for hurricane wind assessment. For example, Gao et al. (2018) presented a new approach for hurricane wind directionality retrieval from SAR images.

Considering the limited documented information on overland inflow angles (Dunn and Miller, 1964; Simpson and Riehl, 1981; Powell, 1982; Powell et al., 1996) a model is proposed based on the available overland inlfow angle references and the model proposed by Zhang and Uhlhorn (2012). Specifically the spatial variability and asymmetry of the over water inflow angle proposed by Zhang and Uhlhorn (2012) is adopted and superimposed with a structure which shifts the inflow angle and further increases it with the radial distance from the center. The over land model used in this paper is given by:

$$\bar{a}_{SR} = a_{SR} + a_{A2}\frac{r^*}{r^*_{max}} + a_{C2} \tag{5.53}$$

where a_{SR} is given by Equation (5.48), a_{C2} is the shifting parameter that is modeled as $a_{C2} \sim N(-15, 2.5)$ and constraint within $[-10, -20]$, and a_{A2} is the scale parameter modeled as $a_{A2} \sim N(-5, 2.5)$ and constraint within $[0, -10]$. Figure 5.34 depicts the streamlines resulting from the simulated wind directionalities in a rectangular mesh for a hurricane prior to landfall. For the depicted streamline realization, overland parameters α_{C2} and α_{A2} were sampled as -14.67° and -4.06°, respectively. The differences on the over ocean and over land inflow angles are clearly distinguished along the coastline.

5.10 Concluding Remarks

The chapter presents a framework for the simulation of synthetic directional wind speeds for hazard and risk analysis. The framework consists of many modules such as the track model, the relative intensity model, the decay model, the radius to maximum wind model, the Holland parameter model, the gradient wind speed model, the boundary layer model and the wind inflow model. Different versions of most of these models are presented for over land and over water cases.

The tropical cyclone track model is based on the Markov Chain approach introduced

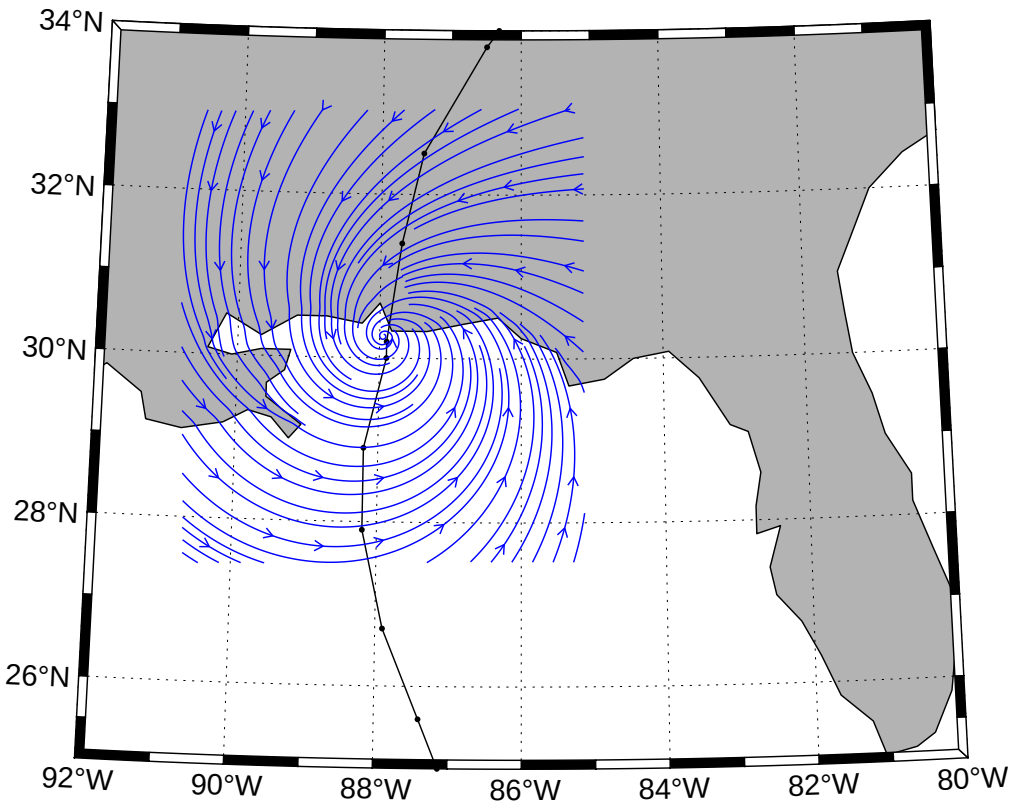


Figure 5.34: Simulated streamlines of the 1-min sustained wind speed at 10 m height. The track and central pressure information is adopted from Hurricane Ivan (2004) at 650 UTC. The sampled RMWis 61.69 km and Holland B parameter is equal to 1.256. The path is depicted with solid black line.

by Emanuel et al. (2006), however, the solution approach herein is different than theirs, in the sense that the Atlantic basin is discretized in a fine mesh and the transition probabilities are developed while considering information from the neighbor cells with weight. Weighting the cells and adjusting the kernel properties can yield good results while keeping the computational time in acceptable levels. In addition, implementation details are presented for all modules and data preprocessing instructions are provided. Assessment of the implemented modules is provided as well, with the constraints of limited available data.

An original overland wind inflow model is proposed. Site-specific wind directionality distributions and joint distributions of wind intensity and directionality can be yielded for any location of the United States East coastline and Gulf of Mexico and used for hazard

analysis or long-term hurricane loss studies. Moreover, estimates of surface wind speed and direction due to a single event can also be extracted for postevent damage assessment.

Chapter 6

Reserach contributions

6.1 Significant contributions

The original contributions of this research are the following:

1. Development of a technique for the optimal representation of multi-dimensional random fields with a moderate number of samples using Functional Quantization. This research showcased the technique's applicability in engineering mechanics involving the simulation of two-dimensional stochastic fields.

2. Development of a methodology named Hazard Quantization for the selection of an optimal set of stochastic intensity measure maps representing the regional hazard over a geographic area, used for the analysis of spatially distributed infrastructure systems. This involved:

    - Showcasing the framework's applicability to seismic hazard
    - Comparing the framework with the current state-of-the-art methodologies in the literature and a benchmark Monte Carlo solution
    - Providing in-depth discussion on the features of each methodology and quantitatively comparing the errors on the hazard curve and auto-correlation
    - Applying Hazard Quantization to hurricane hazard for the optimal selection of wind intensity measure maps

3. Development of a hurricane simulation framework which can model: genesis locations, propagation tracks, size and intensity, key features (e.g., Holland B parameter, regional intensity decay models), wind *intensity* and *directionality*. The developed framework yields probability density functions of hurricane wind directionality and joint probability distributions of the hurricane wind speed and wind directionality at any location of the continental United States which can potentially be exposed to a hurricane. In addition to the site-specific outcomes, the developed framework provides correlated intensity measure maps of maximum wind speed and direction allowing analysts to model the hurricane risk along spatially distributed infrastructure systems.

4. A generalization of the proof that the multi-variate samples generated with a well-known version of the classical Spectral Representation Method (SRM) utilizing discrete Fourier transform (DFT) and frequency double indexing (FDI) are ergodic. The previous proof (Deodatis, 1996) is valid in the limit, when the frequency interval tends to zero and the period tends to infinity; the new proof holds also for the case of finite frequency intervals and samples with finite period, which is the case of practical interest. Moreover, Deodatis (1996) proved that the samples match the *input* CSDM in the limit. In here, it is shown that in the case of practical applications, when the frequency domain is discretized, the *actual* CSDM does not match the *input* CSDM. A closed-form expression of the *actual* resulting CSDM of the generated samples is also derived, and, it is shown that such expression converges to the *input* CSDM in the limit, so generalizing Deodatis (1996) results.

5. A proof demonstrating that the discrete multi-variate samples generated with a well-known version of the classical Spectral Representation Method (SRM) utilizing discrete Fourier transform (DFT) and frequency double indexing technique (FDI) are ergodic, but the resulting spectrum does not match the input (or target) spectrum.

Additionally, a proof demonstrating that the resulting spectrum of the discrete samples matches the input spectrum at the limit as the discretization step approaches zero.

6. Extension an iterative technique which generates one-dimensional, univariate non-Guassian random processes to the case of multi-dimensional univariate non-Gaussian random fields. The random field matches both the arbitrarily prescribed spectral density function and the non-Gaussian marginal distribution.

Appendix A

Passage from continuous to discrete time domain

Let κ be a specific value in the continuous wave number domain: $\kappa \in \mathbb{R}$ and $0 < \kappa \leq \kappa_u$. In the corresponding discretized wave number domain, κ is approximated by $l\Delta\kappa$, where:

$$\Delta\kappa := \kappa_u/N, \quad \text{with} \quad N \in \mathbb{N}^+ \tag{A.1}$$

$$l := \left\lceil \frac{\kappa}{\Delta\kappa} \right\rceil \tag{A.2}$$

which yields that:

$$l \in \mathbb{N} \quad \text{and} \quad 1 \leq l \leq N \tag{A.3}$$

We can observe that when $\Delta\kappa$ tends to 0, then both N and l tend to ∞.

Let's also define the number of components $m \in \mathbb{N}^+$, with $m << N$, and index $j \in \mathbb{N}$ with $1 \leq j \leq m$.

Theorem 1.

$$\lim_{\Delta\kappa \to 0} (l-1)\,\Delta\kappa = \lim_{\Delta\kappa \to 0} l\Delta\kappa = \kappa \tag{A.4}$$

Proof. Due to the ceiling operator in Eq. (A.2), we can bound κ as follows:

$$(l-1)\Delta\kappa < \kappa \leq l\Delta\kappa \tag{A.5}$$

When $\Delta\kappa \to 0$, the bounds of the inequality in Eq. (A.5) tend to coincide, because their

difference tends to 0:

$$\lim_{\Delta\kappa\to 0}\left[l\Delta\kappa-(l-1)\,\Delta\kappa\right]=\lim_{\Delta\kappa\to 0}\Delta\kappa=0 \tag{A.6}$$

Because of the Squeeze Theorem, if the two bounds have the same limit, then also κ has the same limit. In this case we know the latter, so:

$$\lim_{\Delta\kappa\to 0}(l-1)\,\Delta\kappa=\lim_{\Delta\kappa\to 0}l\Delta\kappa=\lim_{\Delta\kappa\to 0}\kappa=\kappa \tag{A.7}$$

□

Corollary 1.

$$\lim_{\Delta\kappa\to 0}\left[(l-1)\,\Delta\kappa+\frac{j}{m}\Delta\kappa\right]=\kappa \tag{A.8}$$

Proof. By definition of j and m, we have that $0<j/m\leq 1$, hence:

$$(l-1)\Delta\kappa<(l-1)\,\Delta\kappa+\frac{j}{m}\Delta\kappa\leq l\Delta\kappa \tag{A.9}$$

When $\Delta\kappa\to 0$, by Theorem 1 and the Squeeze Theorem we obtain the thesis.

□

References

Apivatanagul, P., Davidson, R., Blanton, B., and Nozick, L. [2011]. "Long-term regional hurricane hazard analysis for wind and storm surge." *Coastal Engineering*, 58(6), 499 – 509.

Baker, J. and Jayaram, N. [2008]. "Effects of spatial correlation of ground motion parameters for multi-site seismic risk assessment: collaborative research with stanford university and air." *Report no.*, Final Technical Report. Report for US Geological Survey National Earthquake Hazards Reduction Program (NEHRP) External Research Program Award 07HQGR0031.

Bendat, J. S. and Piersol, A. G. [1986]. *Random Data: Analysis and Measurement Procedures*. Wiley-Interscience, 2 rev exp edition (April).

Benjamin, J. R. and Cornell, C. A. [2014]. *Probability, statistics, and decision for civil engineers*. Courier Corporation.

Blake, E. S., Landsea, C., and Gibney, E. J. [2011]. "The deadliest, costliest, and most intense united states tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts).

Bocchini, P., Christou, V., and Miranda, M. J. [2016]. "Correlated maps for regional multi-hazard analysis: Ideas for a novel approach." *Multi-hazard Approaches to Civil Infrastructure Engineering*, Springer, 15–39.

Bocchini, P., Davison, D., Esnard, M. A., Lamadrid, A., Mitsova-Boneva, D., Sapat, A.,

Sause, R., Snyder, L., and Sun, W. [2019]. "Probabilistic resilience assessment of interdependent systems (praisys), <www.praisys.org>.

Bocchini, P. and Frangopol, D. M. [2011a]. "Generalized bridge network performance analysis with correlation and time-variant reliability." *Structural Safety*, 33(2), 155–164.

Bocchini, P. and Frangopol, D. M. [2011b]. "A stochastic computational framework for the joint transportation network fragility analysis and traffic flow distribution under extreme events." *Probabilistic Engineering Mechanics*, 26(2), 182–193.

Bocchini, P. and Frangopol, D. M. [2012]. "Restoration of bridge networks after an earthquake: multicriteria intervention optimization." *Earthquake Spectra*, 28(2), 426–455.

Bocchini, P., Miranda, M., and Christou, V. [2014]. "Functional quantization of probabilistic life-cycle performance models." *Life-Cycle of Structural Systems: Design, Assessment, Maintenance and Management*, 816–823.

Bommer, J. J. and Crowley, H. [2006]. "The influence of ground-motion variability in earthquake loss modelling." *Bulletin of Earthquake Engineering*, 4(3), 231–248.

Bommer, J. J., Douglas, J., and Strasser, F. O. [2003]. "Style-of-faulting in ground-motion prediction equations." *Bulletin of Earthquake Engineering*, 1(2), 171–203.

Boore, M. D. and Atkison, M. G. [2008]. "Ground-motion prediction equations for the average horizontal component of pga, pgv, and 5and 10.00s." *Earthquake Spectra*, 24(1), 99 – 138.

Boose, E. R., Foster, D. R., and Fluet, M. [1994]. "Hurricane impacts to tropical and temperate forest landscapes." *Ecological Monographs*, 64(4), 369–400.

Brennan, M. J., Hennon, C. C., and Knabb, R. D. [2009]. "The operational use of quikscat ocean surface vector winds at the national hurricane center." *Weather and Forecasting*, 24(3), 621–645.

Brown, S. P. [2006]. "The effect of hurricane katrina on employment and unemployment." *Monthly Lab. Rev.*, 129, 52.

Campbell, K. and Seligson, H. [2003]. "Quantitative method for developing hazard-consistent earthquake scenarios." *Advancing Mitigation Technologies and Disaster Response for Lifeline Systems*, 829–838.

Casson, E. and Coles, S. [2000]. "Simulation and extremal analysis of hurricane events." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(3), 227–245.

CBO [2019]. "Expected costs of damage from hurricane winds and storm-related flooding, <https://www.cbo.gov/system/files/2019-04/55019-ExpectedCostsFromWindStorm.pdf>.

CEUS-SSCn [2012]. "Central and eastern united states seismic source characterization for nuclear facilities: Palo alto, c." *Report no.*, EPRI, and U.S. DOE., and U.S. NRC.

Chang, E. S., Shinozuka, M., and Moore, E. J. [2000]. "Probabilistic earthquake scenarios: Extending risk analysis methodologies to spatially distributed systems." *Earthquake Spectra*, 16(3), 557 – 572.

Chang, S. E. [2003]. "Evaluating disaster mitigations: Methodology for urban infrastructure systems." *Natural Hazards Review*, 4(4), 186–196.

Christou, V. and Bocchini, P. "On the spectrum of multi-variate gaussian, ergodic random samples generated by spectral representation." *(under review)*.

Christou, V. and Bocchini, P. [2014a]. "An efficient methodology that simulates a multi-dimensional non-gaussian field to evaluate the effect of the spatial distribution of corrosion in a steel beam." *Structures Congress 2014*, 1059–1069.

Christou, V. and Bocchini, P. [2014b]. "Optimal representation of multi-dimensional random fields with a moderate number of samples.." *In Computational Stochastic Mechanics 7 (G. Deodatis, P. Spanos eds.)*, CSM7, Santorini, Greece (June 15-18).

Christou, V. and Bocchini, P. [2015]. "Efficient computational models for the optimal representation of correlated regional hazard." *12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12, Vancouver, Canada, July*, Vol. 1215, 18.

Christou, V. and Bocchini, P. [2018]. "Regional hurricane analysis using hazard quantization.." *ASCE Engineering Mechanics Institute Conference, EMI 2018*, Massachusetts Institute of Technology, Cambridge, MA (May 29-Jun 1).

Christou, V., Bocchini, P., and Miranda, J. M. [2016a]. "Optimal approximation of multivariate stochastic processes by functional quantization." *ASCE Engineering Mechanics Institute Conference and Probabilistic Mechanics and Reliability Conference EMI 2016 - PMC 2016*, Vanderbilt University, Nashville, Tennessee, abstract 437 (May 22-25).

Christou, V., Bocchini, P., and Miranda, M. J. [2016b]. "Optimal representation of multi-dimensional random fields with a moderate number of samples: Application to stochastic mechanics." *Probabilistic Engineering Mechanics*, 44, 53–65.

Christou, V., Bocchini, P., Miranda, M. J., and Karamlou, A. [2017]. "Effective sampling of spatially correlated intensity maps using hazard quantization: Application to seismic events." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 4(1), 04017035.

Cimellaro, G. P., Reinhorn, A. M., and Bruneau, M. [2010]. "Framework for analytical quantification of disaster resilience." *Engineering Structures*, 32(11), 3639–3649.

Corlay, S. and Pagès, G. [2015]. "Functional quantization-based stratified sampling methods." *Monte Carlo Methods and Applications*, 21(1), 1–32.

Cornell, C. A. [1968]. "Engineering seismic risk analysis." *Bulletin of the Seismological Society of America*, 58(5), 1583–1606.

Cornell, C. A., Jalayer, F., Hamburger, R. O., and Foutch, D. A. [2002]. "Probabilistic basis for 2000 sac federal emergency management agency steel moment frame guidelines." *Journal of Structural Engineering*, 128(4), 526–533.

Crowley, H. and Bommer, J. J. [2006]. "Modelling seismic hazard in earthquake loss models with spatially distributed exposure." *Bulletin of Earthquake Engineering*, 4, 249–273.

Daniell, J., Khazai, B., Wenzel, F., and Vervaeck, A. [2012]. "The worldwide economic impact of historic earthquakes." *15th WCEE*.

Darling, R. [1991]. "Estimating probabilities of hurricane wind speeds using a large-scale empirical model." *Journal of Climate*, 4(10), 1035–1046.

Decò, A., Bocchini, P., and Frangopol, D. M. [2013]. "A probabilistic approach for the prediction of seismic resilience of bridges." *Earthquake Engineering & Structural Dynamics*, 42(10), 1469–1487.

DeMaria, M., Knaff, J. A., and Kaplan, J. [2006]. "On the decay of tropical cyclone winds crossing narrow landmasses." *Journal of applied meteorology and climatology*, 45(3), 491–499.

Deodatis, G. [1996]. "Simulation of ergodic multivariate stochastic processes." *Journal of Engineering Mechanics*, 122(8), 778–787.

Department of Homeland Security [2003]. *HAZUS-MH 2.1 Earthquake Model Technical Manual.* Department of Homeland Security; Emergency Preparedness and Response Directorate; Federal Emergency Management Agency; Mitigation Division. Whashington, D.C.

Dueñas-Osorio, L. and Vemuru, S. M. [2009]. "Cascading failures in complex infrastructure systems." *Structural safety*, 31(2), 157–167.

Dunn, M. E. and Miller, B. I. [1964]. *Atlantic hurricanes*. Luisiana State University Press, Baton Rouge, Louisiana, USA.

Ebel, J. E. and Kafka, A. L. [1999]. "A monte carlo approach to seismic hazard analysis." *Bulletin of the Seismological Society of America*, 89(4), 854–866.

EIA [2018]. "Hurricane path, not just size, can have a large effect on oil and gas production, <https://www.eia.gov/todayinenergy/detail.php?id=37212>.

Emanuel, K., Ravela, S., Vivant, E., and Risi, C. [2006]. "A statistical deterministic approach to hurricane risk assessment." *Bulletin of the American Meteorological Society*, 87(3), 299–314.

Emanuel, K. A. [1988a]. "The maximum intensity of hurricanes." *Journal of the Atmospheric Sciences*, 45(7), 1143–1155.

Emanuel, K. A. [1988b]. "Toward a general theory of hurricanes." *American Scientist*, 76, 370–379.

Falsone, G. and Impollonia, N. [2002]. "A new approach for the stochastic analysis of finite element modelled structures with uncertain parameters." *Computer Methods in Applied Mechanics and Engineering*, 191(44), 5067–5085.

FEMA, F. E. M. A. [2008]. *Estimated Annualized Earthquake Losses for the United States FEMA 366*. Washington, D.C.

Finance.yahoo [2017]. "What Hurricane Irma could mean for stocks, <https://finance.yahoo.com/news/hurricane-irma-mean-stocks-105038376.html>.

Fiorillo, G., Christou, V., A, S., Bocchini, P., Buceta, J., and Miranda, J. M. [2017]. "Comparison of clustering strategies for the functional quantization of random functions.." *In Safety, Reliability, Risk, Resilience and Sustainability of Structures and Infrastructure (C. Bucher, B.R. Ellingwood, D.M. Frangopol eds.), ICOSSAR, Vienna, Austria, August*, 1327–1336.

Frangopol, M. D., Bocchini, P., Sabatino, S., Christou, V., and Snipes, S. [2015a]. "Probabilistic performance assessment based upon experimental data and optimal planning of maintenance on weapon systems." *Report no.*, Technical Report. Report for the U.S. Department of Defense - Research, Developement, and Engineering Command - Armament Research, Developement and Engineering Center (REDCOM-ARDEC).

Frangopol, M. D., Bocchini, P., Sabatino, S., Christou, V., and Snipes, S. [2015b]. "Probabilistic performance metrics for the assessment and application of maintenance strategies on weapon systems." *Report no.*, Technical Report. Report for the U.S. Department of Defense - Research, Developement, and Engineering Command - Armament Research, Developement and Engineering Center (REDCOM-ARDEC).

Gao, S. and Wang, S. [2018]. "Progressive collapse analysis of latticed telecommunication towers under wind loads." *Advances in Civil Engineering*, 2018.

Gao, Y., Guan, C., Sun, J., and Xie, L. [2018]. "A new hurricane wind direction retrieval method for sar images without hurricane eye." *Journal of Atmospheric and Oceanic Technology*, 35(11), 2229–2239.

Gardoni, P., Mosalam, K. M., and der Kiureghian, A. [2003]. "Probabilistic seismic demand models and fragility estimates for rc bridges." *Journal of Earthquake Engineering*, 7(spec01), 79–106.

Georgiou, P., Davenport, A. G., and Vickery, B. [1983]. "Design wind speeds in regions dominated by tropical cyclones." *Journal of Wind Engineering and Industrial Aerodynamics*, 13(1-3), 139–152.

Georgiou, P. N. [1986]. "Design wind speeds in tropical cyclone-prone regions." Ph.D. thesis, The University of Western Ontario, The University of Western Ontario.

Gersho, A. and Gray, R. M. [1991]. *Vector quantization and signal compression*. Kluwer Academic Publishers, Norwell, MA, USA.

Goda, K. and Atkinson, G. M. [2009]. "Probabilistic characterization of spatially correlated response spectra for earthquakes in japan." *Bulletin of the Seismological Society of America*, 99(5), 3003–3020.

Goda, K. and Hong, H. [2008a]. "Estimation of seismic loss for spatially distributed buildings." *Earthquake Spectra*, 24(4), 889–910.

Goda, K. and Hong, H. [2008b]. "Spatial correlation of peak ground motions and response spectra." *Bulletin of the Seismological Society of America*, 98(1), 354–365.

Goda, K. and Hong, H. [2009]. "Deaggregation of seismic loss of spatially distributed buildings." *Bulletin of Earthquake Engineering*, 7(1), 255–272.

Goklany, I. M. [2009]. "Deaths and death rates from extreme weather events: 1900-2008." *Global Trends*, 13, 14.

Graham, H. E. and Nunn, D. E. [1959]. "Meteorological considerations pertinent to standard project hurricane, atlantic and gulf coasts of the united states." *Report no.*, U.S. Department of Commerce, Weather Bureau, Washington, D.C. Report No. 33.

Grigoriu, M. [1995]. "Applied non-gaussian processes: Examples, theory, simulation, linear random vibration, and matlab solutions(book)." *Englewood Cliffs, NJ: Prentice Hall, Inc, 1995.*

Grigoriu, M. [2006]. "Evaluation of karhunen–loève, spectral, and sampling representations for stochastic processes." *Journal of engineering mechanics*, 132(2), 179–189.

Grigoriu, M. [2009]. "Reduced order models for random functions. application to stochastic problems." *Applied Mathematical Modelling*, 33(1), 161–175.

Grigoriu, M. [2010]. "Linear random vibration by stochastic reduced-order models." *International journal for numerical methods in engineering*, 82(12), 1537–1559.

Grigoriu, M. [2013]. "Solution of linear dynamic systems with uncertain properties by stochastic reduced order models." *Probabilistic Engineering Mechanics*, 34, 168–176.

Hall, T. M. and Jewson, S. [2007]. "Statistical modelling of north atlantic tropical cyclone tracks." *Tellus A: Dynamic Meteorology and Oceanography*, 59(4), 486–498.

Han, Y. and Davidson, R. A. [2012]. "Probabilistic seismic hazard analysis for spatially distributed infrastructure." *Earthquake Engng. Struct. Dyn*, 41, 2141–2158.

Holland, G. J. [1980]. "An analytic model of the wind and pressure profiles in hurricanes." *Monthly weather review*, 108(8), 1212–1218.

James, M. and Mason, L. [2005]. "Synthetic tropical cyclone database." *Journal of waterway, port, coastal, and ocean engineering*, 131(4), 181–192.

Jarvinen, B. R., Neumann, C. J., and Davis, M. A. [1984]. "A tropical cyclone data tape for the north atlantic basin, 1886-1983: Contents, limitations, and uses." *Report no.*, U.S. Department of Commerce, Washington, D.C. NOAA Tech. Memo. NWS-NHC-22.

Jayaram, N. and Baker, J. W. [2009]. "Correlation model for spatially distributed ground-motion intensities." *Earthquake Engineering & Structural Dynamics*, 38(15), 1687–1708.

Jayaram, N. and Baker, J. W. [2010]. "Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment." *Earthquake Engng. Struct. Dyn*, 39, 1109–1131.

Ju, L., Du, Q., and Gunzburger, M. [2002]. "Probabilistic methods for centroidal voronoi tessellations and their parallel implementations." *Parallel Computing*, 28(10), 1477–1500.

Kaplan, J. and DeMaria, M. [2001]. "On the decay of tropical cyclone winds after landfall in the new england area." *Journal of applied meteorology*, 40(2), 280–286.

Karamlou, A. and Bocchini, P. [2014]. "Optimal bridge restoration sequence for resilient transportation networks." *Structures Congress 2014*, ASCE, 1437–1447.

Karamlou, A. and Bocchini, P. [2016]. "Sequencing algorithm with multiple-input genetic operators: application to disaster resilience." *Engineering Strucutres*, 117, 591–602.

Karamlou, A., Bocchini, P., and Christou, V. [2016]. "Metrics and algorithm for optimal retrofit strategy of resilient transportation networks." *Maintenance, Monitoring, Safety, Risk and Resilience of Bridges and Bridge Networks*, Foz do Iguau, Brazil, June 26–30, Taylor and Francis, 1121–1128.

Kepert, J. [2001]. "The dynamics of boundary layer jets within the tropical cyclone core. part i: Linear theory." *Journal of the Atmospheric Sciences*, 58(17), 2469–2484.

Kiremidjian, S. A., Stergiou, E., and Lee, R. [2007]. "Issues in seismic risk assessment of transportation networks." *Earthquake Geotechnical Engineering*, K. Pitilakis, D., ed., Springer Netherlands, 461–480.

Kramer, S. L. [1996]. *Geotechnical earthquake engineering*, Vol. 80. Prentice Hall Upper Saddle River, NJ.

Kunz, M., Mühr, B., Kunz-Plapp, T., Daniell, J., Khazai, B., Wenzel, F., Vannieuwenhuyse, M., Comes, T., Elmer, F., Schröter, K., et al. [2013]. "Investigation of superstorm sandy 2012 in a multi-disciplinary approach." *Natural Hazards and Earth System Sciences*, 13(10), 2579–2598.

Lee, R. and Kiremidjian, S. A. [2007]. "Uncertainty and correlation for loss assessment of spatially distributed systems." *Earthquake Spectra*, 23(4), 753–770.

Legg, R. M., Nozick, k. L., and Davidson, A. R. [2010]. "Optimizing the selection of hazard-consistent probabilistic scenarios for long-term regional hurricane loss estimation." *Structural Safety*, 32(1), 90 – 100.

Lejay, A. and Reutenauer, V. [2012]. "A variance reduction technique using a quantized brownian motion as a control variate." *The Journal of Computational Finance*, 16(2), 61–84.

Liu, F. [2014]. "Projections of future us design wind speeds due to climate change for estimating hurricane losses.

Lloyd, S. [1982]. "Least squares quantization in pcm." *Information Theory, IEEE Transactions on*, 28(2), 129–137.

Luschgy, H. and Pagès, G. [2002]. "Functional quantization of gaussian processes." *Journal of Functional Analysis*, 196(2), 486–531.

Luschgy, H., Pagès, G., et al. [2004]. "Sharp asymptotics of the functional quantization problem for gaussian processes." *The Annals of Probability*, 32(2), 1574–1599.

Luschgy, H., Pagès, G., and Wilbertz, B. [2010]. "Asymptotically optimal quantization schemes for gaussian processes on hilbert spaces." *ESAIM: Probability and Statistics*, 14, 93–116.

Ma, L., Bocchini, P., and Christou, V. "Fragility models of electrical conductors in power transmission networks subjected to hurricanes." *Structural safety, in press*.

Ma, L., Bocchini, P., and Christou, V. [in press 2019]. "Fragility models of electrical conductors in power transmission networks subjected to hurricanes.

MacQueen, J. [1967]. "Some methods for classification and analysis of multivariate observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley, Calif., University of California Press, 281–297.

Manzour, H., Davidson, R. A., Horspool, N., and Nozick, L. K. [2016]. "Seismic hazard and loss analysis for spatially distributed infrastructure in christchurch, new zealand." *Earthquake Spectra*, 32(2), 697–712.

Mara, T. and Hong, H. [2013]. "Effect of wind direction on the response and capacity surface of a transmission tower." *Engineering structures*, 57, 493–501.

MATLAB [2010]. *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts.

McGuire, R. K. [1976]. "Fortran computer program for seismic risk analysis." *Report no.*, US Geological Survey,.

McGuire, R. K. [2004]. *Seismic hazard and risk analysis*. Earthquake engineering research institute.

McGuire, R. K. [2008]. "Probabilistic seismic hazard analysis: Early history." *Earthquake Engineering & Structural Dynamics*, 37(3), 329–338.

Mignolet, M. P. and Soize, C. [2008]. "Stochastic reduced order models for uncertain geometrically nonlinear dynamical systems." *Computer Methods in Applied Mechanics and Engineering*, 197(45-48), 3951–3963.

Miranda, M. and Bocchini, P. [2013]. "Functional quantization of stationary gaussian and non-gaussian random processes." *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*, 2785–2792.

Miranda, M. J. and Bocchini, P. [2015a]. "A versatile technique for the optimal approximation of random processes by functional quantization." *Applied Mathematics and Computation*, 271, 935–958.

Miranda, M. J. and Bocchini, P. [2015b]. "A versatile technique for the optimal approximation of random processes by functional quantization." *Applied Mathematics and Computation*, 271, 935–958.

Musson, R. [1999]. "Determination of design earthquakes in seismic hazard analysis through monte carlo simulation." *Journal of Earthquake Engineering*, 3(04), 463–474.

Musson, R. [2000]. "The use of monte carlo simulations for seismic hazard assessment in the uk." *Annals of Geophysics*, 43(1).

Neumann, C. J. [1991]. "The national hurricane center risk analysis program(hurisk).

NHC [2014]. "Atlantic hurricane database (HURDAT2), <https://www.aoml.noaa.gov/hrd/hurdat/newhurdat-format.pdf>.

NHC/NOAA [2018]. "Coastliest U.S. tropical cyclone tables updated, <https://www.nhc.noaa.gov/news/UpdatedCostliest.pdf>. Accessed: 2019-06-03.

NOAA [2017]. "Socioeconomic Data Summary, <https://coast.noaa.gov/data/digitalcoast/pdf/socioeconomic-data-summary.pdf>.

Padgett, J. E., DesRoches, R., and Nilsson, E. [2010]. "Regional seismic risk assessment of bridge network in charleston, south carolina." *Journal of Earthquake Engineering*, 14(6), 918–933.

Pagès, G. and Printems, J. [2005]. "Functional quantization for numerics with an application to option pricing." *Monte Carlo Methods and Applications mcma*, 11(4), 407–446.

Papadopoulos, V. and Papadrakakis, M. [2005]. "The effect of material and thickness variability on the buckling load of shells with random initial imperfections." *Computer Methods in Applied Mechanics and Engineering*, 194(12-16), 1405–1426.

Park, J., Bazzurro, P., Baker, J., et al. [2007]. "Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation." *Applications of statistics and probability in civil engineering. Taylor & Francis Group, London*, 1–8.

Peterson, D. M., Moschetti, P. M., Powers, M. P., Mueller, S. C., Haller, M., K., Frankel, D. A., Yuehua, Z., Sanaz, R., Harmsen, C. S., Oliver, B. S., Field, N., Chen, R., Rukstales, S. K., Luco, N., Wheeler, L. R., Williams, A. R., and Olsen, H., A. [2014]. "Documentation for the 2014 update of the united states national seismic hazard maps." *Open-file report 2014-1091*, USGS.

Popescu, R., Deodatis, G., and Nobahar, A. [2005]. "Effects of random heterogeneity of soil properties on bearing capacity." *Probabilistic Engineering Mechanics*, 20(4), 324–341.

Porter, K. A. [2003]. "An overview of peers performance-based earthquake engineering methodology." *Proceedings of Ninth International Conference on Applications of Statistics and Probability in Civil Engineering*.

Powell, M. [2008]. "High wind drag coefficient and sea surface roughness in shallow water." *NOAA/AOML Hurricane Research Division Final Rep. to the Joint Hurricane Testbed*.

Powell, M., Soukup, G., Cocke, S., Gulati, S., Morisseau-Leroy, N., Hamid, S., Dorst,

N., and Axe, L. [2005]. "State of florida hurricane loss projection model: Atmospheric science component." *Journal of wind engineering and industrial aerodynamics*, 93(8), 651–674.

Powell, M. D. [1982]. "The transition of the hurricane frederic boundary-layer wind field from the open gulf of mexico to landfall." *Monthly Weather Review*, 110(12), 1912–1932.

Powell, M. D., Houston, S. H., and Reinhold, T. A. [1996]. "Hurricane andrew's landfall in south florida. part i: Standardizing measurements for documentation of surface wind fields." *Weather and Forecasting*, 11(3), 304–328.

Powell, M. D., Uhlhorn, E. W., and Kepert, J. D. [2009]. "Estimating maximum surface winds from hurricane reconnaissance measurements." *Weather and Forecasting*, 24(3), 868–883.

Rhoades, D. A. and McVerry, G. H. [2001]. "Joint hazard of earthquake shaking at two or more locations." *Earthquake Spectra*, 17(4), 697–710.

Ruiz-Garc´ia, J. and Miranda, E. [2007]. "Probabilistic estimation of maximum inelastic displacement demands for performance-based design." *Earthquake engineering & structural dynamics*, 36(9), 1235–1254.

Russell, L. R. [1969]. "Probability distributions for texas gulf coast hurricane effects of engineering interest.." Ph.D. thesis, Stanford University, Stanford University.

Sanchez-Sesma, J., Aguirre, J., and Sen, M. [1988]. "Simple modeling procedure for estimation of cyclonic wind speeds." *Journal of Structural Engineering*, 114(2), 352–370.

Saydam, D., Bocchini, P., and Frangopol, D. M. [2013]. "Time-dependent risk associated with deterioration of highway bridge networks." *Engineering Structures*, 54, 221–233.

Schwerdt, R. W., Ho, F. P., and Watkins, R. R. [1979]. "Meteorological criteria for standard project hurricane and probable maximum hurricane windfields, gulf and east coasts of the united states.

Shen, H., He, Y., and Perrie, W. [2009]. "Speed ambiguity in hurricane wind retrieval from sar imagery." *International Journal of Remote Sensing*, 30(11), 2827–2836.

Shields, M., Deodatis, G., and Bocchini, P. [2011]. "A simple and efficient methodology to approximate a general non-gaussian stationary stochastic process by a translation process." *Probabilistic Engineering Mechanics*, 26(4), 511–519.

Shinozuka, M. and Deodatis, G. [1991]. "Simulation of stochastic processes by spectral representation." *Applied Mechanics Reviews*, 44(4), 191–204.

Shinozuka, M. and Deodatis, G. [1996]. "Simulation of multi-dimensional gaussian stochastic fields by spectral representation." *Applied Mechanics Reviews*, 49(1), 29–53.

Shinozuka, M. and Jan, C. M. [1972]. "Digital simulation of random processes and its applications." *Journal of Sound and Vibration*, 25(1), 111–128.

Simpson, R. H. and Riehl, H. [1981]. *The hurricane and its impact*. Luisiana State University Press, Baton Rouge, Louisiana, USA.

Sokolov, V. and Wenzel, F. [2011a]. "Influence of ground-motion correlation on probabilistic assessments of seismic hazard and loss: sensitivity analysis." *Bulletin of Earthquake Engineering*, 9(5), 1339–1360.

Sokolov, V. and Wenzel, F. [2011b]. "Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation." *Earthquake Engineering & Structural Dynamics*, 40(9), 993–1009.

Stefanou, G. and Papadrakakis, M. [2007]. "Assessment of spectral representation and karhunen–loève expansion methods for the simulation of gaussian stochastic fields." *Computer methods in applied mechanics and engineering*, 196(21-24), 2465–2477.

Takewaki, I. [2011]. "Preliminary report of the 2011 off the pacific coast of tohoku earthquake." *Journal of Zhejiang University SCIENCE A*, 12(5), 327–334.

Teixeira, A. P. and Soares, C. G. [2008]. "Ultimate strength of plates with random fields of corrosion." *Structure and Infrastructure Engineering*, 4(5), 363–370.

Tesfamariam, S. and Goda, K. [2013]. *Handbook of seismic risk analysis and management of civil infrastructure systems*. Elsevier.

Thompson, E. F. and Cardone, V. J. [1996]. "Practical modeling of hurricane surface wind fields." *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 122(4), 195–205.

Vaziri, P., Davidson, R., Apivatanagul, P., and Nozick, L. [2012]. "Identification of optimization-based probabilistic earthquake scenarios for regional loss estimation." *J. of Earthquake Engineering*, 16, 296–315.

Version, A. [2013]. "6.13, analysis users manual." *Dassault Systemes Simulia Corp., Providence, RI*.

Vickery, P., Skerlj, P., and Twisdale, L. [2000a]. "Simulation of hurricane risk in the us using empirical track model." *Journal of structural engineering*, 126(10), 1222–1237.

Vickery, P. J. [2005]. "Simple empirical models for estimating the increase in the central pressure of tropical cyclones after landfall along the coastline of the united states." *Journal of applied meteorology*, 44(12), 1807–1826.

Vickery, P. J., Skerlj, P., Steckley, A., and Twisdale, L. [2000b]. "Hurricane wind field model for use in hurricane simulations." *Journal of Structural Engineering*, 126(10), 1203–1221.

Vickery, P. J. and Twisdale, L. A. [1995]. "Prediction of hurricane wind speeds in the united states." *Journal of Structural Engineering*, 121(11), 1691–1699.

Vickery, P. J. and Wadhera, D. [2008]. "Statistical models of holland pressure profile parameter and radius to maximum winds of hurricanes from flight-level pressure and h* wind data." *Journal of Applied Meteorology and climatology*, 47(10), 2497–2517.

Vickery, P. J., Wadhera, D., Powell, M. D., and Chen, Y. [2009a]. "A hurricane boundary layer and wind field model for use in engineering applications." *Journal of Applied Meteorology and Climatology*, 48(2), 381–405.

Vickery, P. J., Wadhera, D., Twisdale Jr, L. A., and Lavelle, F. M. [2009b]. "Us hurricane wind speed risk and uncertainty." *Journal of structural engineering*, 135(3), 301–320.

Vogl, S. [2009]. "Tropical cyclone boundary-layer models." Ph.D. thesis, Technical University of Munich, Technical University of Munich.

Wang, M. and Takada, T. [2005]. "Macrospatial correlation model of seismic ground motions." *Earthquake spectra*, 21(4), 1137–1156.

Warner, J. E., Grigoriu, M., and Aquino, W. [2013]. "Stochastic reduced order models for random vectors: Application to random eigenvalue problems." *Probabilistic Engineering Mechanics*, 31, 1–11.

Wells, D. L. and Coppersmith, K. J. [1994]. "New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement." *Bulletin of the seismological Society of America*, 84(4), 974–1002.

Wesson, R. L. and Perkins, D. M. [2001]. "Spatial correlation of probabilistic earthquake ground motion and loss." *Bulletin of the Seismological Society of America*, 91(6), 1498–1515.

Zerva, A. [1992]. "Seismic ground motion simulations from a class of spatial variability models." *Earthquake Engineering & Structural Dynamics*, 21(4), 351–361.

Zhang, J. A. [2019]. "Lack of over land inflow studies (April). private communication.

Zhang, J. A. and Uhlhorn, E. W. [2012]. "Hurricane sea surface inflow angle and an observation-based parametric model." *Monthly Weather Review*, 140(11), 3587–3605.

Vita

Vasileios Christou was born in Ioannina, Greece in 1984. He received his Diploma in Civil Engineering from Aristotle University of Thessaloniki in 2009. In 2011, he completed his M.Sc. degree in Structural Engineering from University of Patras. In his Masters thesis he worked on the mathematical modeling of multiple concave sliding bearings and the numerical solution of the non-linear equations of motions when these isolation devices are applied to structures and excited by near fault earthquake ground motions. Vasileios joined Lehigh University in 2012 to continue his education towards his Ph.D. He will receive his Ph.D. degree in August 2019.